National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

MORAL INCONTINENCE IN GAUTHIER'S MORALS BY AGREEMENT

BY

© KENN F.T. CUST

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS.

DEPARTMENT OF PHILOSOPHY

EDMONTON, ALBERTA

FALL 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-45820-8

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Kenn F.T. Cust

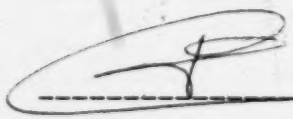
TITLE OF THESIS: Moral Incontinence in Gauthier's Morals by Agreement

DEGREE: Master of Arts

YEAR THIS DEGREE GRANTED: 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.



(Student's signature)

6011-84 Ave

(Student's permanent address)

Edmonton, Alberta T6B-0H3

DATE: July 7/88

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for Acceptance, a thesis entitled Moral Incontinence in Gauthier's Morals By Agreement submitted by Kenn F.T. Cust in partial fulfillment of the requirements for the degree of Master of Arts.

.....
(Supervisor)

.....

.....

.....

Date: June 24, 1988

Abstract

David Gauthier in his recent work, Morals by Agreements, argues, amongst other things, that given the choice between choosing a constrained maximizing disposition (a CM disposition) and a straightforwardly maximizing disposition (an SM disposition) a rational agent, conceived of as a rational utility maximizer, would choose the former disposition rather than the latter. Moreover, he argues that the constrained maximizing disposition is the uniquely rational strategy given the two alternatives.

My response to Gauthier's argument in favour of constrained maximization being the uniquely rational strategy is fourfold. First I argue that a CM disposition and an SM disposition are not jointly exhaustive dispositions. I then identify a third disposition, a C-S disposition, and distinguish this third disposition from Gauthier's two alternatives on behavioral and disposition grounds. I then argue that there are cases under which it would be more rational, in terms of utility maximization, to choose a C-S disposition rather than a CM disposition. I also argue that a C-S disposition is a more plausible disposition for a rational agent, conceived of as a rational utility maximizer, to hold.

The implication of my arguments for Gauthier's constrained maximizers is that they are morally incontinent: constrained

maximizers are moral when they need not be whereas C-S's merely take the occasional "moral holiday." I offer support for the conclusion that Gauthier's CM is morally incontinent by arguing that the conditions which a CM takes to be necessary and sufficient for him to base his actions on a joint strategy are sometimes only necessary conditions for a C-S to base her actions on a perceived joint strategy and at other times these conditions are both necessary and sufficient for her to base her actions on an actual joint strategy.

Table of Contents

CHAPTER		PAGE
	Introduction	1
I.	THE GAUTHIER MORAL ENTERPRISE	10
	The Morally Free Zone	12
	Minimax Relative Concession	16
	The Revised Lockean Proviso	23
	Constrained Maximization	29
	The Archimedean Point	38
II.	THE MOTIVATIONAL COMPONENT	45
III.	THE MORALLY INCONTINENT CM	87
	Bibliography	117
	Appendix A: The Technical Foundation	119
	Appendix B: Calculated Solutions	137

Introduction

We live in a world with other people; we are constantly making agreements with them, agreements which are, for the most part, mutually beneficial. When entering into cooperative ventures with our fellow man we expect both parties to the agreement to comply with what they have agreed upon. If either party fails to abide by the terms and conditions of the agreement, then we seek redress through the appropriate channels. Suppose there were no legal channels, suppose there were no political institutions and no sovereign governing our daily social intercourse with others, then what? Would we still make agreements? And if so, what would motivate us to comply with those agreements, once made? Consider these questions from the eyes of a fully rational agent, where rationality is given the instrumental definition of "utility maximization."

David Gauthier, in his recent book, Morals By Agreement, argues that not only would a utility maximizer make agreements, he would also comply with those agreements, even given the absence of legal channels, political institutions and sovereigns.

His argument, which rests on certain idealizing assumptions and an assumption about what it means to be rational, i.e., utility maximization, is that it would be rational for agents so defined to enter and keep agreements. Of course, this is not meant to imply that we should make agreements, and keep agreements, with everyone whom we meet. On the contrary, Gauthier is quite specific; he argues that a certain type of rational agent, broadly conceived of as a constrained utility maximizer, with coherent and considered preferences, will find it to his individual advantage if he makes agreements with others, and keeps those agreements, insofar as those agreements maximize his utility in terms of his coherent and considered preferences.

Gauthier's instrumental conception of rationality as utility maximization is the foundation of his overall argument. Building on this foundation he attempts to demonstrate how impartial constraints on the pursuit of individual utility maximizing behaviours can be generated. His claim is that he can generate these impartial constraints without importing into his argument any prior moral assumptions. To support this end he distinguishes a disposition of constrained maximization of utility (a CM disposition) from a disposition of straightforward maximization of utility (an SM disposition). Then he argues that it would be more rational to choose a disposition of constrained maximization of utility rather than a disposition of straightforward maximization of utility. Given these two alternatives, Gauthier

argues, the disposition of constrained maximization of utility is the uniquely rational strategy.

Gauthier introduces Hobbes's Foole to play the role of a rational agent who adopts a straightforward utility maximizing strategy and raises the problem of compliance.

> Is not the Foole's ultimate argument that the truly prudent person, the fully rational utility maximizer, must seek to appear trustworthy, an upholder of agreements? For then he will not be excluded from the cooperative arrangements of his fellows, but will be welcomed as a partner, while he awaits opportunities to benefit at their expense -- and, preferably, without their knowledge, so that he may retain the guise of constraint and trustworthiness.[1]

The classic, formal, statement of the compliance problem is found in the Prisoner's Dilemma (PD). I reiterate Gauthier's description of the tale.

> Fred and Ed have committed (the District Attorney is certain) a serious crime, but some of the evidence necessary to secure a conviction is, unfortunately, inadmissable in court (the District Attorney curses the law reformers who make her task more difficult). She is, however, holding Fred and Ed, and has been able to prevent them from being in a position to communicate one with the other. She has them booked on a lesser, although still serious, charge, and she is confident that she can secure their conviction for it. She then calls, separately, on Fred and Ed, telling each the same tale and making each the same offer. 'Confess the error of your ways -- and the crime you have committed,' she says, 'and if your former partner does not confess, then I shall convince the jury that you

---

[1] Gauthier, David; Morals By Agreement; Clarendon Press, Oxford; 1986, p. 173.

> are a reformed man and your ex-partner evil incarnate; the judge will sentence you to a year and him to ten. Do not confess, and if your former partner does, then you may infer your fate. And should neither of you choose to confess, then I shall bring you to trial on this other matter, and you may count on two years.' 'But what,' says Fred (or Ed), 'if we both confess?' 'Then,' says the District Attorney, 'I shall let justice take its natural course with you -- it's a serious crime so I should estimate five years,' and without further ado she leaves Ed (or Fred) to solitary reflection. [2]

The problem for Fred and Ed is extremely serious if they are straightforward utility maximizers. If Fred and Ed both confess then each faces five years in jail; if neither Fred nor Ed confess they each face two years in jail. This is the best solution given the description of the problem. If Fred confesses, and Ed does not, then Fred faces only one year in jail while Ed faces ten years in jail. If Ed confesses, and Fred does not, then Ed faces only one year in jail while Fred would face ten years in jail. I now return to Gauthier's description of the reasoning involved for both Fred and Ed.

> Both are quite single-mindedly interested in minimizing their time behind bars, and so both represent the situation in this simple way:

	He confesses	He does not confess
I confess	5 years each	1 year for me,

[2] Morals By Agreement; op cit., pp 79-80. This dilemma is credited to A. W. Tucker and is recounted by Luce, Duncan R. and Raiffa, Howard, in Games and Decisions: Introduction and Critical Survey; John Wiley & Sons, Inc., New York, 1958, p. 94.

		10 for him
do not confess	10 years for me, 1 for him	2 years each

So each reasons: 'If he confesses, then I had better confess -- otherwise I'm in for ten years. If he doesn't confess, then if I confess I'm out in a year. So whatever he does, I should confess. Indeed' (each being a student of strategic rationality and the theory of games) 'in this situation there is but one outcome in equilibrium -- the product of mutual confession. My course is clear.'

And the District Attorney (also a student of strategic rationality and the theory of games) returns to Fred and Ed on the following day, a smile on her face, to collect the confessions which she knows are forthcoming, and which will put Fred and Ed out of evil paths for five years -- whereas, if they had kept their mouths shut (or their pens dry), they would have been back at their favorite occupations in but two years. [3]

Straightforward utility maximizers, even if they agreed not to confess should they ever find themselves in a Prisoner's Dilemma situation, would confess if they were strictly concerned with their own interest. They would both confess, because neither could be certain that the other would not confess.

The Prisoner's Dilemma has great theoretical importance for it encapsulates the problem of strategically rational choice in cases of strategic interaction: basically, each actor, in such a two-person one-shot Prisoner's Dilemma, must base his choice of action on his expectation of the other's choice of action. Formally there are three conditions imposed on strategically

---

[3] ibid.

rational choices.

> A: Each person's choice must be a rational response to the choices she expects the others to make.
>
> B: Each person must expect every other person's choices to satisfy condition A.
>
> C: Each person must believe her choice and expectations to be reflected in the expectations of every other person.[4]

In the Prisoner's Dilemma faced by Fred and Ed these three conditions of strategically rational choice dictate Fred and Ed's solution to the problem they find themselves in. Each confesses not only because neither can rely on the other not to confess but also because their reasoning reflects these three conditions of strategically rational choice.

Gauthier argues that if rational agents were to choose a disposition of constrained maximization of utility, rather than a

---

[4] Morals by Agreement, op cit., p. 61.

> Condition A relates the rationality of the actor to the framework of interaction; it requires her to be strategically rational. Condition B makes explicit the assumption that all parties to the interaction are strategically rational and that this rationality is a matter of common knowledge. Condition C makes explicit the assumption that each person views the situation as if her knowledge of the grounds of choice were complete, shared by all, and known by all to be so shared.

disposition of straightforward maximization of utility, then, even given the three formal conditions of strategically rational choice, his constrained maximizers of utility can expect to do better in a two-person one-shot Prisoner's Dilemma than could straight-forward maximizers of utility.

Of course there are some further assumptions Gauthier makes, the virtue of translucency being an important one, and his argument is far more complex than the brief synopsis I have offered here. What is important about Gauthier's argument is that he challenges the received view of what utility maximization consists of, and in doing so he is better equipped to answer some of the objections which have long plagued various philosophies of egoism. Gauthier goes so far as to suggest that his argument for constrained maximization can solve the formal problem of the Prisoner's Dilemma.

If Gauthier's argument is as good as he claims many of our current social systems and structures would have to be reexamined. Political solutions might no longer be required to resolve the problems posed by agents acting strictly to maximize their own utilities; social programs would require new and innovative justifications. Compelled by the logic of Gauthier's reasoning we would have to accept his conclusions on a new moral order.

However, Gauthier's moral enterprise is not successful. One mistake he makes is to assume that a disposition of constrained maximization of utility and a disposition of straightforward maximization of utility are jointly exhaustive dispositions. I argue that this is not the case by identifying a third alternative, a C-S disposition. I argue that this third alternative disposition can readily be distinguished from either of Gauthier's two alternatives, if not on behavioral grounds then on dispositional grounds. I also argue that this C-S disposition is a more plausible conception of what an agent, conceived of as a utility maximizer, would be like. And moreover I argue that this C-S disposition is at least as rational as Gauthier's CM disposition in some cases. That is to say that I can find conditions under which it would be rational, in terms of utility maximization, to dispose oneself to a C-S disposition. The implication for Gauthier, assuming my argument to be valid, is first that Gauthier's constrained maximizer of utility is morally incontinent; that is to say he is moral when the dictates of rationality do not require him to be. The second implication is that Gauthier's CM disposition is not the uniquely rational strategy he thinks it is.

In putting forth my argument for this C-S disposition I have been careful to present it on terms which do not violate any of Gauthier's assumptions. Furthermore, I have had to identify and clarify some issues which Gauthier was not too clear on; I then

used these clarifications to my advantage when doing so was permitted.

The first chapter of this paper consists of an overview of Gauthier's project. I identify the five central ideas of his overall argument and present them to the reader for the purpose of being able to relate the major argumentative portion of my thesis to Gauthier's overall task. Chapter II and III are the argumentative portions of this thesis. In these two chapters I set out in detail my argument for a C-S disposition and demonstrate why it is at least as rational a strategy in some cases as Gauthier's CM disposition. There are two appendixes to this thesis; the first offers an explication of Gauthier's technical apparatus while the second appendix gives the results of various calculations to validate my argument. Though my argument is essentially critical of Gauthier it can be viewed as a constructively critical thesis. I think Gauthier can use and build on some of the arguments I have presented to strengthen his position.

Chapter I

The Gauthier Moral Enterprise

The goal of David Gauthier's recent book Morals By Agreement is to argue for a "contractarian rationale for morality."[1] In order to provide a contractarian rationale for morality Gauthier must accomplish two separate tasks. First he must define the content of his principles of morality, then he must provide reasons for complying with these content-laden moral principles. Gauthier is aware of both of these tasks and, in setting forth to fulfill them, he offers an argument that will challenge philosophers for decades to come.

In a Hobbesian state of nature [2] rational agents make no distinction between what they may and may not do, what they should and should not do. Gauthier argues that insofar as it can

---

[1] Morals By Agreement; op cit., p. 9. This thesis has profited immensely from discussions with Professor Wes Cooper, Professor Steven DeHaven, and graduate students Jill Hunter and Joe Mackenzie.

[2] A "state of nature" can basically be understood as a pre-government state of humanity, where the descriptions fall anywhere between the extremes offered by Hobbes and Rousseau. In Hobbes' description we have a "war of all against all" where life is "nasty, brutish and short." Rousseau's description of the state of nature is somewhat more positive; in his description we have the "noble savage" who lives a life completely opposite to the one described by Hobbes.

For Gauthier's project he has in mind the Hobbesian portrait of the state of nature. If we were all noble savages living the life Rousseau depicts, morals by agreement would be completely unnecessary.

be shown that these agents will accept constraints on the pursuit of their individual utility [3] maximizing behaviors then these constraints can be viewed as moral principles.

> We shall argue that the rational principles for making choices, or decisions among possible actions, include some that constrain the actor pursuing his own interest in an impartial way. These we identify as moral principles.[4]

Thus, for Gauthier, moral principles consist of rationally self-imposed constraints. This is the "substantive component" of his moral theory. His "motivational component" will be illustrated shortly. [5]

---

[3] Gauthier initially defines utility as a measure of preference, (where preferences have to be both considered and coherent), but, to be more precise, he subsequently defines utility as "a measure of outcomes representing relations of preference." See Morals By Agreement op cit., p. 23. To say that preferences are considered is to say

> there is no conflict between their behavioral and attitudinal dimensions and they are stable under experience and reflection.

To say that preferences are coherent is to say that they meet the formal conditions of game theory for ordinal and interval measures. This is discussed in Appendix A.

Kurt Baier, in a yet unpublished paper, given at the Morals By Agreement Conference at Bowling Green State University, in April 1987, entitled "Rationality, Value and Preference," identifies four different conceptions of preference and raises problems for Gauthier's instrumental conception of rationality as the maximization of utility, i.e., the measure of coherent and considered preferences.

[4] Morals By Agreement; op cit., p. 3.

[5] In "Morality and the Theory of Rational Choice;" Ethics; Volume 97, Number 4, July 1987, University of Chicago; pp. 715-716, by Kraus, Jody S., and Coleman, Jules L., they make the

There are five central components which provide the foundation for morals by agreement: a morally free zone, minimax relative concession (MRC), constrained maximization, a revised Lockean Proviso, and a Archimedean Point. Using these components and an idea borrowed from economic theory, rationality defined as utility maximization, Gauthier begins to construct an edifice that many, if not all Libertarians, would be proud to inhabit. Libertarians would want to inhabit Gauthier's moral world for they would find in it all the freedom they would want in terms of interacting with one another plus the absence of coercion normally found in political answers to the problems of strategic interaction. I will examine each of these central components in turn, demonstrating their importance for the edifice Gauthier wants to construct.

The Morally Free Zone

The morally free zone proves to be a beast which inhabits economic theory, the perfectly competitive market. [6] In a

---

distinction between the substantive and the motivational components of a contractarian moral theory.

> The substantive component specifies the content of the principles of morality; the motivational component explains why a rational person would comply with the principles specified by the substantive theory.

[6] The perfectly competitive market can be described as one where there is private ownership of the means of production, where there is division of labour and a market exchange of goods and services. It also means that there is absolutely no interference in the market by any external factors, including the

perfectly competitive market there is no need for morality--that is, for impartial constraints on the pursuit of individual utility maximization -- because the perfectly competitive market ensures that each individual can pursue his activities without costs to another. In the perfectly competitive market mutual advantage, Pareto-optimality, [7] is guaranteed; what I do, in a perfectly competitive market, in no way affects the optimality of the outcome for you. This is the case even though we are not concerned with each other's interest; this is Gauthier's "non-tuism." [8]

However, in the real world of ships, sealing wax, cabbages

---

government. It is also generally assumed that there is no cost involved in the exchange of goods and services. In Gauthier's terms, a perfectly competitive market presupposes that there are no externalities, no free riders and no parasites. See Morals By Agreement, pp. 83-100. For further details about the perfectly competitive market see von Mises, Ludwig, Human Action; Contemporary Books, Inc., Chicago, 1963, pp. 237-239.

[7] To say that an outcome is a "Pareto-optimal" outcome is just to say that "there is no possible outcome affording some person a greater utility and no person a lesser utility." See Morals By Agreement; p. 76.

[8] Morals By Agreement; op cit., p. 87. "The market requires only that persons be conceived as not taking an interest in the interests of those with whom they exchange." Christopher Morris, in a yet unpublished paper entitled "The Relation Between Self-Interest and Justice in Contractarian Ethics," presented at the Morals By Agreement Conference in April 1987 distinguishes between "asociality, egoism, mutual unconcern, non-tuism, and materialism" and argues that it isn't always clear which of these conditions Gauthier is relying on at different stages of his argument." (See Morris, p. 10) While this may be an interesting observation on Morris' part it will not be directly relevant to my present concerns.

and kings, perfectly competitive markets do not obtain. [9] There are economic inefficiencies, externality problems, free-riders and parasites,[10] hence a need for morality. The outcomes in the real world are not Pareto-optimal, i.e., there is some possible outcome affording some person a greater utility and no person a lesser utility. What Gauthier wants to argue is that morality,

---

[9] One must be vigilant, when moving from the idealizing assumptions made for the perfectly competitive market to the real world of partial compliance, that these idealizing assumptions are not carried over into the real world of ships, sealing wax, cabbages and kings. In partial compliance theory we drop the idealizing assumptions which have been made and consider the world, and the people in it, as one would most likely find it.

> It (partial compliance theory) includes, among other things, the theory of punishment and compensatory justice, just war and conscientious objection, civil disobedience and militant resistance.*

* See John Rawls, A Theory of Justice; The Belnap Press of Harvard University Press, Cambridge, MA; 1971; p. 351.

[10] Morals By Agreement; op cit., p. 87. There are two kinds of externality problems, positive externalities and negative externalities.

> An externality arises whenever an act of production or exchange or consumption affects the utility of some person who is not party, or who is unwilling party, to it. Such an effect may of course be either beneficial or harmful; if beneficial we speak of a positive externality or external efficiency, if harmful we speak of a negative externality or external inefficiency.

Corresponding with positive externalities are free riders and corresponding with negative externalities are parasites.

> A free-rider obtains a benefit without paying all or part of the cost. A parasite in obtaining a benefit displaces all or part of the cost on some other person.

The fundamental distinction between free-riders and parasites is that the free-rider does not make others directly worse off whereas the parasite does.

impartial constraints on the pursuit of individual utility maximization, is a possible solution, and a better solution than a political solution, to the problems of market failure.

A political solution to solve the problems of market failure would involve some substantive rules of conduct, rules which would generate specific conduct in a given situation. The agents involved would not be free to act in any manner whatsoever as they would be allowed to given a moral solution. In contrast with a political solution, a moral solution allows the rational agents to adopt some procedural rule such that it maximizes their utility while mitigating the need for them to be concerned with the needs and interests of others. In a political solution, rules are made and the agents follow them whether or not their own utilities are maximized, thus obtaining, if not a Pareto-optimal solution then a Pareto-efficient solution.

By identifying a need for morality through the heuristic device of the perfectly competitive market Gauthier also opened the proverbial Pandora's Box thereby providing me with the means to plague Gauthier's conception of a constrained utility maximizer. I will show in the forthcoming chapters that adopting a constrained maximizing disposition is not the uniquely rational strategy that Gauthier thinks it is. If there is a need for morality, and Gauthier argues that there is, then it must be the case that the possibility for acting immorally exists. By

identifying this possibility I am able to argue that given sufficiently rare opportunities for acting immorally a rational agent might be better off if he adopted the alternative of a C-S disposition rather than the constrained maximizing disposition which Gauthier favours.

Minimax Relative Concession

Gauthier's approach for showing that morality is a possible solution to market failure is to use a rational bargaining theory. In order to ensure that a rational bargaining approach would be effective in grounding morality in rationality, Gauthier has to show basically two things: first that it is rational to agree to cooperate and second that it is rational to comply with those agreements and constrain one's pursuit of individual utility maximization.

The rational bargaining approach for resolving market failures can be broken into two separate problems. The first problem is the outcome of the bargaining itself. One must be able to arrive at an outcome which is Pareto-optimal. The second problem involved with using a rational bargaining approach for solving market problems is the initial bargaining situation itself. With what are the bargainers allowed to approach the table?

The outline for Gauthier's bargaining apparatus begins with some common assumptions about rationality and the agents involved in game theoretic approaches to bargaining. Rationality, as I have said, is initially defined as utility maximization.

> Let it further be agreed that in so far as the interests of others are not affected, a person acts rationally if and only if she seeks her greatest interest or benefit. [11]

Gauthier goes on to make some other idealizing assumptions about rational agents being fully informed, equally rational, and translucent to one another.

While these assumptions are innocent enough, in that they are assumptions which are commonly made in game theory, they still pose a problem for strategic interaction. [12] The main problem they raise for Gauthier, and anyone else who attempts to walk on these waters, is that it must be shown how it can be the case that the parties to an interaction can maximize their utilities given the three conditions imposed on strategically

---

[11] Morals By Agreement; op cit., pp. 6-7. In the course of setting out the argument for morals by agreement this conception of rationality as utility maximization on the level of particular choices will be reinterpreted such that rationality will then be identified with utility-maximization at the level of dispositions to choose. See Morals By Agreement pp. 182 - 184.

[12] Morals By Agreement; op cit., p. 21. Strategic interaction is just when an actor takes his behaviour to be but one variable among others. This is opposed to parametric interaction where the actor takes his behaviour to be the sole variable in a fixed environment.

rational choice. This is the formal problem illustrated by Fred and Ed when they found themselves in the Prisoner's Dilemma.

Gauthier's solution to this first problem is his idea of "minimax relative concession" (MRC). We can understand minimax relative concession to mean that in the cooperative enterprise, which results from the bargaining process, all the co-operative surplus, (the joint sum realized by the cooperative venture minus each participants factor costs), is divided equally amongst the two participants of the joint venture. [13] Gauthier's technical definition is somewhat more complicated. He says that

> in any cooperative interaction, the rational joint strategy is determined by a bargain among the cooperators in which each advances his maximal claim and then offers a concession no greater in relative magnitude than the minimax concession.[14]

Gauthier argues that the equal rationality of his utility maximizers will lead them to accept MRC as the principle one would appeal to both in terms of the process of agreement and the content of that agreement.[15]

---

[13] Morals By Agreement; op cit., p. 145.

[14] Morals By Agreement; op cit., p. 145.

[15] Gauthier assumes that his agents are equally rational and fully informed. These are not uncommon assumptions within the framework of game theoretic approaches to bargaining and generally they are considered to be unproblematic provided that one stays within the boundaries of ideal theory. See Games and Decisions: Introduction and Critical Survey; op cit., p. 5.

This (i.e., the agent being fully informed about

To illustrate the application of MRC consider the case of Sam McGee the prospector and Grasp the Banker.[16] Sam McGee has discovered a very significant gold reserve in the Yukon but he lacks the necessary cash, say $100.00, to register his claim. McGee approaches Grasp the Banker to secure the necessary funds in order to register his claim. There is no one else that McGee can approach for funds (Grasp is the only banker and no one else has the resources to lend McGee the funds). Given that McGee and Grasp the Banker are both rational utility maximizers a question arises as to how the co-operative surplus will be distributed, assuming that they reach an agreement about cooperating.

---

another's utility functions*), and the kindred assumption about his ability to perceive the game situation, are often subsumed under the phrase "the theory assumes rational players." Though it is not apparent from some writings, the term "rational" is far from precise,** and it certainly means different things in the different theories that have been developed. Loosely, it seems to include any assumption one makes about the players maximizing something, and any about complete knowledge on the part of the player in a very complex situation, where experience indicates that a human being would be far more restrictive in his perceptions.

* A "utility function" is simply the numerical representation of a person's preferences if those preferences are consistent in a given prescribed manner. See Luce and Raiffa, p. 4.

** Joseph Mendola, in a recent article in Ethics, Vol. 97, No. 4, July 1987, argues that Gauthier fails to distinguish between two senses of rationality, "rationality as intelligibility" and "rationality as justification."

16 Morals By Agreement; op cit., pp. 153-154.

McGee can't register his claim without the assistance of Grasp the Banker and therefore, if he cannot obtain the funds necessary to register the claim, he can expect no return on his labours. Grasp the Banker, on the other hand, has not assisted McGee in any way with respect to finding the claim; McGee has just approached him for the money. Grasp, having read Morals By Agreement, and being familiar with the principle of MRC, tells McGee he will loan him the money to register the claim if McGee will give him a half-share in the claim, i.e., a 50% share of the co-operative surplus. Rationally, Gauthier argues, poor McGee will have to accept the banker's demands. "For although Grasp's $100.00 is worth only $100.00 in the absence of McGee's discovery, yet McGee's discovery is worthless to him without Grasp's money." [17]

Minimax relative concession is appealed to by both Grasp the Banker and McGee the Prospector during the process of agreement and it is appealed to as part of the content of their agreement. Both reason that if there is no agreement then McGee's claim will be worth nothing; there will be no co-operative surplus. However, if there is agreement then the co-operative surplus will need to be fairly distributed. MRC is the principle appealed to for distributing this co-operative surplus.

To satisfy the rational bargaining conditions, Gauthier

---

[17] Morals By Agreement; op cit., p. 153.

argues, McGee and Grasp, as equally rational agents, go through the following procedure. First, both McGee and Grasp advance their maximal claim; each claims the entire co-operative surplus. Since neither would accept the maximal claim the other has made, if they want to reach an agreement they then have to minimize their maximum relative concession. That is to say each reasons that there is a "feasible concession point" that every rational person is willing to entertain and they minimize their own maximum concession relative to this feasible concession point. Once they have reasoned to minimizing their maximum relative concession they can then reach an agreement. As equally rational and fully informed agents, Gauthier argues, McGee and Grasp both stand to gain from social cooperation; therefore they each would minimize their maximum relative concession and reach agreement.

In bargaining situations such as this, the "willingness to concede" to the feasible concession point, Gauthier says "expresses the equal rationality" of the bargainers. [18] This is to be distinguished from, though it doesn't conflict with, the instrumental conception of rationality as utility maximization in those cases where straightforward maximizers, (SM's), and constrained maximizers of utility, (CM's), reason parametrically over a joint or an individual strategy, or those cases where a rational agent reasons parametrically over a constrained maximizing or a straightforward maximizing disposition. This

---

[18] Morals By Agreement; op cit. p. 143.

distinction becomes clearer as I proceed.

James Fishkin, in a yet unpublished paper, challenges Gauthier's principle of minimax relative concession, MRC. Fishkin argues, using the example cited below, that there may be cases where one does not stand to gain any surplus from social cooperation. Hence, there may be no rational reason, on Gauthier's own terms, for such individuals to enter into society. MRC will not provide a sufficient reason for social cooperation even given the nonviolation of Gauthier's revised Lockean Proviso.

Claude Brown, Fishkin says, conducted a study on some violent youths in the Harlem Ghetto. "It included several accounts of how their violent behaviour seemed rational. Brown challenged them with the following reasoning about how a life of crime was not in their interest."

> Every time he goes on the prowl, for a victim or an establishment, he runs the risk of one of three serious misfortunes: there is a 60% chance he will be killed, permanently maimed or end up doing a long bit in jail. And even if he succeeds in getting over nine or ten times for $1,000 or more, in a few days to a week at most he will be right back where he started -- at the bottom of the hill... Couldn't he see how futile it was?

"Brown was startled by the response to this challenge by a youth he describes as not yet old enough to shave but who was doing a fifteen year sentence for armed robbery:"

> "I see where you comin' from, Mr. Brown" he replied, "but you got things kind of turned around the wrong way. You see, all the things you say could happen to me is dead on the money and that is why I can't lose. Look at it from my point of view for a minute. Let's say I go and get wiped (killed). Then I ain't got no more needs, right? All my problems are solved. I don't need no more money, no more nothing, right? OK, supposin' I get popped, shot in the spine and paralyzed for the rest of my life -- that could happen playing football, you know. Then I won't need a whole lot of money because I won't be able to go no place and do nothing, right? Now if I get busted and end up in the joint pulling a dime and a nickel like I am (15 year sentence) then I don't have to worry about no bucks, no clothes. I get free rent and three square meals a day. So you see Mr. Brown, I don't really lose." [19]

Given the reasoning of this child from the ghetto, there is no rational reason for him to end his life on the edge of society and make an agreement to enter society for society cannot, he believes, offer him anything worthwhile to make such an agreement. He is in such a poor position to begin with that there would be no co-operative surplus, if he did enter society, for minimax relative concession (MRC) to be appealing to him.

The Revised Lockean Proviso

However, prior to even considering entering an agreement one must first find acceptable what each potential participant brings

---

[19] Fiskin, James; "Bargaining, Justice and Justification: Towards Reconstruction," presented at the Morals By Agreement Conference in Bowling Green, Ohio, April 1987, pp. 7-8. Fiskin cites as his source, Claude Brown, "Manchild in Harlem" New York Times Magazine, Sept. 17, 1984, p. 44.

to the table.

> If persons are willing to comply with the agreement that determines what each takes from the bargaining table, then they must find initially acceptable what each brings to the table. And if what some bring to the table includes the fruit of prior interaction forced on their fellows, then this initial acceptability will be lacking. If you seize the products of my labour and then say 'Let's make a deal,' I may be compelled to accept, but I will not voluntarily comply. [20]

Gauthier's solution to this second problem is the revised Lockean Proviso. The revised Lockean Proviso "prohibits bettering one's situation through interaction that worsens the situation of another." [21] In other words, if we are to engage in cooperative interaction, neither of us must have been disadvantaged by the efforts of the person with whom we are considering cooperating. If either of us has been so disadvantaged then, for any cooperative arrangement, compliance would not be ensured and

---

[20] Morals By Agreement; op cit., p. 15.

[21] Morals By Agreement; op cit., p. 205. Also see p. 200, 201. This leaves open, of course, the possibility that I might have violated the Lockean Proviso with respect to some other person, someone that I am not interested in bargaining with. Having done so, if I now approach you in an attempt to bargain, and you have no knowledge of my previous violation of the Proviso, then you would be at a strategic disadvantage. Gauthier does not attempt to deal with this problem and in actual fact he invites the raising of this problem when he says (pp. 200-201)

> We shall argue that the terms of fully rational co-operation include the requirement that each individual's endowment, affording him a base utility not included in the co-operative surplus, must be considered to have been initially acquired by him without taking advantage of any other person -- or, more precisely, of any other cooperator.

social stability would be questionable.

Consider the example Gauthier uses to illustrate that there is a distinction to be made between worsening another's situation and failing to better it. There are two individuals, P and Q, and Q is in the water and about to drown. [22] For our purposes we will consider two ways in which this situation could have arisen. In the first case Q could have been standing, innocently, on the bank and P could have pushed Q in. In the second case, Q could have fallen into the water by accident. The two possible outcomes we want to consider are the outcomes where P saves Q and where P does not save Q.

In the first case it could be said that P worsened Q's situation, for he pushed Q in the water. In the second case it could be said that P did not worsen Q's situation, for he had nothing to do with Q's being in the water.

One might want to argue that in the second case P did, in fact, worsen Q's situation; P could have saved Q, even though P had nothing to do with Q's being in the water, and by failing to do so he makes Q worse off than Q need be. However, intuitive

---

[22] Morals By Agreement; op cit., p. 204. The example also occurs in Robert Nozick's essay "Coercion" in Philosophy, Politics and Society, Fourth Series; edited by Peter Laslett, W.G. Runciman and Quentin Skinner; p. 115. This example is also discussed, in relation to Gauthier's Morals By Agreement by James S. Fiskin in a paper, entitled "Bargaining, Justice and Justification: Towards Reconstruction," op cit.

though this argument may be, it misses the point Gauthier wants to establish. Gauthier wants to use this example as an analogy for the initial bargaining situation such that he can determine a "base point" from which bargaining may proceed.

The agreement made by Grasp the Banker and Sam McGee may not be in accord with our considered intuitions, as is P's failing to save Q, but that is no reason for rejecting Gauthier's argument. Gauthier warns us several times that

> No doubt there will be differences, perhaps significant, between the impartial and rational constraints supported by our argument and the morality learned from parents and peers, priests and teachers. [23]

We take heed of this warning and caution the reader that we cannot reject Gauthier's argument just because we don't like some of the implications. Gauthier may respond, as H.P. Grice is reported to have responded on one occasion, "See here, that's not an objection to my theory -- that's my theory!"[24]

The essential difference, for Gauthier, between these two

---

[23] Morals By Agreement; op cit., p. 6. Also see page 179 where he says "First, we should not suppose that our argument upholds all of conventional morality, or all of those institutions and practices that purport to realize fair and optimal outcomes."

[24] Rachels, James; "Feeding the Hungry: Killing and Starving to Death" in Moral Issues, edited by Jan Narveson; Oxford University Press, Canada, 1983; p. 163. Rachels reports that Grice made this remark at a conference in Oberlin "several years ago."

cases is that in the first case where P pushes Q in the water, Q is made worse off; but in the second case it is just the same as if P had never arrived on the scene. According to Gauthier Q cannot say that P made him, Q, worse off because if P had not arrived on this tragic scene, and therefore did not save Q, then this is the same as P arriving on the scene, (not having pushed Q in the water to begin with), and not saving Q. Gauthier's base point for determining whether someone is made better or worse off, (and what each can approach the bargaining table with) is determined by what one can expect in the absence of another. In the absence of P, Q can expect to drown.

One might want to argue, as Fishkin does, that if we applied MRC to this example of P and Q then Q must, rationally, on Gauthier's terms, agree to give P half of his future earnings. By P's saving Q, a co-operative surplus is created, such that, as rational agents concerned to further their own interests, Q must accede to P's demands for 50% of his future earnings, and "surely this is not essential justice." Gauthier can counter this objection by pointing out that neither P nor Q would have objected to this situation when considering it from an ideal perspective [25] which mitigated their knowledge of who either might be in this particular case. Neither P nor Q, when considering this possible situation from this ideal perspective,

---

[25] This ideal perspective is Gauthier's conception of an Ideal Actor choosing from a Archimedean Point and it will be explained in more detail as I proceed.

would have known who they would have been in this unfortunate example, and therefore they would have agreed to the outcome governed by MRC, as long as the Proviso had not been violated.

Gauthier acknowledges the role the Proviso plays in his theory in his Preface to the text. To place the quote in its proper context, Gauthier was discussing three central problems he had to overcome during the many years he spent working on the theory. Concerning the third problem, the role of the Proviso, he says,

> And the third was to determine the appropriate initial position from which cooperation proceeds, which requires showing the rationality of accepting a Lockean Proviso on initial acquisition. (...) This third problem proved to be the most recalcitrant; from the initial idea of a contractarian moral theory, which captured my imagination in 1966, some thirteen years elapsed before the role of the proviso became clear.[26]

Rational agents would only consider approaching the table if they knew that what each initially brought to the table had been acquired fairly; that is to say that neither of the rational agents would have been placed at a strategic disadvantage by the coercive efforts of the other. If this initial acquisition was unfair then the bargaining situation itself would be contaminated such that any expected outcome would be unfair; this would lead to problems with compliance and hence social instability.

---

[26] Morals By Agreement; op cit., p. v.

The Proviso, Gauthier says, "is not the product of rational agreement. Rather, it is a condition that must be accepted by each person for such agreement to be possible." [27] Should the potential participants to the agreement fail to accept some initial position such that the bargaining process could not get off the ground then, while Gauthier might have shown that there was a need for morality with his morally free zone, that would be the extent of his accomplishment.

Constrained Maximization

Assuming that Gauthier is successful in showing why rational utility maximizers might make agreements then, in order to supply the motivational component of his theory, he must show that these rational utility maximizers would comply with the agreements they have made. Why should Q pay P, after P has saved him? As a rational utility maximizer, isn't it in Q's best interest to defect from the agreement once he is on terra firma?

One straightforward reason, which was aptly phrased by Hobbes, is that life would be "nasty, brutish and short" if men didn't find some way, political or moral, to restrain themselves. A second reason why rational utility maximizers might be concerned to keep the contracts they have made is that they might stand to benefit more from cooperating with others. However,

---

[27] Morals By Agreement; op cit., p. 16.

while they might stand to gain more from entering into agreement it doesn't follow that they would benefit more from keeping the agreement rather than defecting from those agreements when it was in their interest to do so.

In fact, one might surely argue, the rational , i.e., the utility maximizing thing to do, is to be a party to any cooperative arrangement where one would stand to gain more from being a party to the arrangement than not being a party to the arrangement. But in those cases where defecting from the arrangement yields a greater expected utility than following through with the arrangement then the utility maximizing strategy is to defect. Q might simply say to P, "your demands for half of my future wealth is a price too dear for me to pay. I will not comply with the terms of our agreement." Gauthier has to provide Q with the reasons for complying with his agreement, a difficult task if there ever was one.

To solve the compliance problem Gauthier makes a distinction between those actors who are straightforward maximizers of utility (SM's) and those who are constrained maximizers of utility (CM's). An SM adopts a disposition to maximize her utility in any particular situation that she finds herself in, whereas a CM decides to adopt a disposition such that she will refrain from straightforwardly maximizing her utility if she truly believes herself to be amongst others who share this

disposition and her expected utility is nearly fair and optimal. The important distinction between these two dispositions is that the straightforward maximizer takes into account the strategies of those with whom he interacts while the constrained maximizer takes into account the utilities of those with whom he interacts.

> Let us say that a straightforward maximizer is a person who seeks to maximize his utility given the strategies of those with whom he interacts. A constrained maximizer, on the other hand, is a person who seeks in some situations to maximize her utility, given not the strategies but the utilities of those with whom she interacts. [28]

This move to solve the compliance problem is crucial for Gauthier and requires some elaboration. Gauthier begins with rational agents in a Hobbesian state of nature. He makes the idealizing assumptions that these rational agents are both equally rational and fully informed. He then argues that a rational individual would choose, given the alternatives of either adopting an SM disposition or a CM disposition, to adopt constrained maximization as his disposition for strategic

---

[28] Morals By Agreement; op cit., p. 167. I will argue that one can identify a further disposition, (I call it a C-S disposition), and argue that it is at least as rational a strategy, in some cases, as Gauthier's constrained maximizing disposition. That is to say, given certain assumptions, I can find conditions under which it would be rational to choose a C-S disposition rather than a CM disposition.

interaction. [29]

It is important to note that Gauthier makes a move that others involved with game theory have not made. The accepted interpretation amongst game theorists is that they identify rationality at the level of particular choices. Thus on the received view a choice would be rational if and only if (iff) it maximized the actor's expected utility on that particular occasion. But on Gauthier's view choice involves dispositions. He says,

> We identify rationality with utility maximization at the level of dispositions to choose. A disposition is rational if and only if an actor holding it can expect his choices to yield no less utility than the choices he would make were he to hold any alternative disposition. [30]

This move is important, Gauthier argues, for it holds the key to answering Hobbes's Foole and it also provides the motivational component his theory needs, thus (allegedly) solving the

---

[29] Gauthier makes a distinction between "strategic choice" and "parametric choice." In parametric choice "the actor takes his behaviour to be the sole variable in a fixed environment" whereas in strategic choice, which is actually a case of interaction, the "actor takes his behaviour to be but one variable among others, so that his choice must be responsive to his expectations of others' choices, while their choices are similarly responsive to their expectations." This is covered in more detail in Appendix A but for now the reader can refer to Morals By Agreement p. 21.

[30] Morals By Agreement; op cit., pp. 182-183.

compliance problem.[31]

In order for morals by agreement to be a viable alternative to a political solution Gauthier has to show that rational agents in a Hobbesian state of nature (or the prisoners in the Prisoner's Dilemma) will come to accept impartial constraints on the pursuit of their utility maximizing behaviors such that they will comply with the agreements that they have made. If Gauthier fails to demonstrate that rational agents will voluntarily impose these constraints on themselves then the Gauthier enterprise will have failed (and the poor unfortunate prisoners will have no recourse but to spend the next five years in jail).

Gauthier argues that adopting either a CM disposition or an SM disposition affects the situations one could reasonably expect to find oneself in. He then argues that CM's will fare better than SM's.

> A straightforward maximizer, who is disposed to make maximizing choices, must expect to be excluded from cooperative arrangements which he would find advantageous. A constrained maximizer may expect to be included in such arrangements. She benefits from her disposition, not in the choices she makes, but in her opportunities to choose.[32]

---

[31] I will take advantage of this move on Gauthier's part when I construct my argument in favour of a C-S disposition.

[32] Morals By Agreement; op cit., p. 183. I will argue that a person who chooses a C-S disposition has this advantage over the CM disposition and that she will ultimately fare better in terms of overall utility maximization than would a person with a CM

There are some additional characteristics of CM's that Gauthier argues for in support of his argument for compliance. First, CM's (and SM's) are "translucent" as opposed to being either "transparent" or "opaque." To say that persons are translucent is just to say that we can ascertain their disposition to cooperate or not cooperate, "not with certainty, but as more than mere guesswork." [33] Gauthier rejects transparency in favour of translucency because his argument in favour of a constrained maximizing disposition would have little practical import in the real world of ships, sealing wax, cabbages and kings if he only managed to show this given the idealizing assumption that all persons were transparent. He thinks that persons really are, to some degree or other, translucent and this helps to port his argument over to the real world.

Gauthier also argues that CM's would be "narrowly compliant" as opposed to being "broadly compliant." A narrowly compliant person is one "who is disposed to cooperate in ways that, followed by all, yield nearly optimal and fair outcomes," and a person who is broadly compliant is a person who is disposed to cooperate in ways that, followed by all, "merely yield her some

---

disposition.

[33] Morals By Agreement; op cit., p. 174. I will have some things to say about Gauthier's conception of translucency in Chapters II and III.

benefit in relation to universal non-compliance." [34]

There is another important aspect to Gauthier's argument and that involves the choice to be an SM or a CM. Gauthier repeatedly emphasizes that a choice is involved when a rational agent is considering whether to adopt a CM or an SM disposition. Gauthier makes little mention of what is involved in being able to choose this disposition, of what is involved after one makes this choice, or what is involved in choosing a disposition to be narrowly compliant.

Kraus and Coleman raise this issue in a footnote of their article but then drop it. They indicate that they would have some conceptual difficulties in realizing what such dispositions could be like but for the purposes of their argument they need not consider the issue and allow Gauthier to "help himself to such a disposition." [35] I am sympathetic to both of these author's

---

[34] Morals By Agreement; op cit., p. 178. In the article by Kraus and Coleman, op cit, they argue, and argue very forcefully, that Gauthier fails to show that narrow compliance is uniquely rational. I will not consider the issue of narrow compliance for this would detract from my main concern; however, these authors have made a very forceful case against Gauthier in this regard.

[35] See Kraus and Coleman, op cit., p. 722.

> The disposition being considered, it would seem can be no ordinary one. Rather, it would appear to be more akin to a self-imposed psychological guarantee of compliance, one which cannot be resisted. If it could be resisted, then it would appear that straightforward maximizers would resist it in PD-structured bargains in

concerns but nevertheless while the problems they have identified need addressing I think we can, for the sake of argument, offer a response on Gauthier's behalf. [36]

Choosing a disposition is somewhat like choosing a general principle which will govern one's actions over a wide range of cases. This is not to say that the chosen principle will prove to be adequate for all cases for it will not. However it will prove adequate for most cases and therefore the principle is sufficient to motivate us to act such that we need not spend all our time

---

> order to maximize utility, and so adoption of it would not result in successful collective actions. But because Gauthier does want to claim that rational actors so disposed are still making a choice when they cooperate in the PD, he wants the disposition to fall short of psychological compulsion; yet, because he wants it to influence their choice to cooperate, it must be strong enough to override consideration of direct utility maximization. It is difficult to conceptualize what such a disposition could be and even more troublesome to contemplate its psychological plausibility. Nonetheless, for present purposes, we allow Gauthier to help himself to such a disposition.

[36] One question about dispositions which one might want to consider concerns the issue of factor rent. Gauthier argues that a complete tax on factor rent cannot affect the optimality of the outcome. (See pp. 272-277.) However, using Gauthier's example of Gretzky, assuming that Gretzky would play hockey for less than the current salary he receives and where factor rent is the "difference between the least amount that would induce him to play as well as he does and his actual remuneration," then one might want to consider why Gretzky would not change his disposition so that there was no difference between the least amount that would induce him to play as well as he does and his actual remuneration. I think an argument could be made that it would be rational for Gretzky to alter his disposition so that there was no factor rent and I think that such an argument could be supported by appealing to two of Gauthier's ideas, i.e., liberal man and free affectivity.

deciding whether or not to act in any particular case. If we had to make a decision concerning each particular case, weighing all the implications, considering all the alternatives and so on, we would be completely immobilized and never make any decisions. On the other hand, by adopting some general principle, some rule of thumb that will guide our actions through the most trivial and generally important cases, we save ourselves the trouble of questioning our general principle in all but the extreme cases. While this response may not be sufficient to assuage the concerns of either Kraus or Coleman I think it can provide the reader with a sufficient grasp of what Gauthier means when he says rational agents adopt a disposition thereby letting the argument proceed.

There is however another problem with Gauthier's conception of dispositions which is not addressed by Kraus and Coleman. This problem is epistemological in nature and can be traced back historically at least to Hobbes, Hume, and Kant. In each of these authors' writings they found the need for a political solution to the problems of market failure on the grounds that actors might be mistaken in their perceptions and/or their apprehensions of what was going on around them. Gauthier fails to note that there is a distinction to be made between acting on the circumstances one finds one's self in and acting on the perception and/or apprehension of the circumstances one finds one's self in. In any given situation agents can make epistemological errors and therefore the need for a solution, be it a political or moral

solution, to the problems of market failure is not economic inefficiencies but rather epistemological deficiencies.

Returning to the Prisoner's Dilemma: if both prisoners were CM's, who were both narrowly compliant and reasonably translucent, (and who hopefully had read Gauthier's Morals By Agreement), then we can see how they could obtain the Pareto-optimal outcome, i.e., only serving two years. They would have made an agreement, such that if they ever found themselves in a situation analogous to the Prisoner's Dilemma, then they would not confess. Because they were CM's, who were narrowly compliant and reasonably translucent, and not SM's, they could expect that the other would not confess by virtue of their each having chosen the disposition to be a constrained maximizer. If they were both SM's then they could not, reasonably, have this expectation.

The Archimedean Point

The last central idea of the Gauthier moral enterprise is that of the Ideal Actor in the Archimedean Point. The Archimedean Point is somewhat similar to the role Rawls' "original position"[37] plays in his Theory of Justice. The Ideal Actor's role is analogous to the role the POP's (people in the original position) play. The function of Gauthier's Ideal Actor in the Archimedean Point is to choose among principles of interaction. In concrete

---

[37] A Theory of Justice; op cit., p. 12.

terms, the Ideal Actor makes a choice "among social structures embodying these principles of interaction." [38]

Gauthier makes some assumptions about the Ideal Actor and imposes some conditions on her. First, we are to consider the Ideal Actor to be fully rational and fully informed. We are also to consider the Ideal Actor to suffer from certain ignorance conditions such that she cannot "identify herself as a particular person within society." [39] The Ideal Actor is also considered to be a utility maximizer. Thus equipped she is ready to choose the principles of social interaction and the social structures which will embody those principles.

These idealizing assumptions give the Ideal Actor the information and reasoning capability necessary to make the choices she is to make. The ignorance condition ensures impartiality on the part of the Ideal Actor. If she cannot identify herself as any particular person within society, Gauthier argues, then her choices will be acceptable to all members of society, (assuming that they are equally rational and fully informed as she is). The Ideal Actor, Gauthier argues, must choose "not as if she had an equal chance of being each of the persons affected by her choice," (this is what Rawls' POP's base

---

[38] Morals By Agreement; op cit., pp. 233-234.

[39] Morals By Agreement; op cit., pp. 235-236.

their decision on), "but as if she were each of those persons." [40]

Gauthier also raises the issue of personal identity with respect to the Ideal Actor choosing from the Archimedean Point:

> The self at time t_1 is identical with the later self at time t_2 to the extent that it identifies with that later self, and this identification is measured by the weight given to the expected preferences of the self at t_2 in the preferences of the self at t_1. [41]

Given this description of personal identity, if the Ideal Actor must choose as if she were each of those persons who would be affected by her choice then this choice will accommodate the possibility of one's future self such that some future self will be able to identify and accept that choice, given the ignorance conditions under which the Ideal Actor operates under.

What choices, given these conditions and assumptions, would the Ideal Actor in the Archimedean Point make? The answer to this question should not surprise us. Gauthier argues that she would choose minimax relative concession, the revised Lockean Proviso, constrained maximization, and the free market to embody the principles of interaction which would govern the new moral world.

---

[40] Morals By Agreement; op cit., p. 255.

[41] Morals By Agreement; op cit., p. 343.

This is the essence of Gauthier's Morals by Agreement, with some of the particular details omitted. Gauthier has been described as "coming closer to pulling off this 'argument of a lifetime' than anyone could legitimately expect," [42] and I share these sentiments. However, while he may have come closer than anyone expected to pulling it off, his argument for the motivational component of his theory does suffer from a very serious problem; it doesn't work.

The dispositions of constrained and straightforward maximization are not jointly exhaustive dispositions. I will argue in the following chapters, amongst other things, that one can identify a third disposition, distinct from either of Gauthier's alternatives, and that this third alternative is at least as rational, in some cases, as Gauthier's CM disposition. For the moment let us contrast Gauthier's CM and SM dispositions with this third disposition, call it a C-S disposition. In the following chapters I will provide much more detail on this third disposition but for now I merely want to highlight, through a concrete example, what one could expect from a rational agent who chose one of these three alternative dispositions.

Consider Rosencrantz, Guildenstern and Becassine [43] where

---

[42] Kraus and Coleman, op cit., p. 721.

[43] "The Relation Between Self-Interest and Justice In Contractarian Ethics," op cit., p. 47. Becassine is a character introduced by Morris in an example he uses to illustrate that

Rosencrantz has adopted a CM disposition, where Guildenstern has adopted an SM disposition and where Becassine has adopted a C-S disposition. For the purposes of this example we consider them to be translucent such that their disposition to cooperate can be reasonably ascertained, "not with certainty but as more than mere guesswork." [44]

Rosencrantz, having adopted a CM disposition, encounters Guildenstern, an SM, who wants her to assist him with bringing in his crop. Given translucency Rosencrantz is able to determine with some degree of accuracy that Guildenstern is an SM. Rosencrantz's response is to act straightforwardly, as is Guildenstern's, therefore they do not cooperate. We can even imagine Rosencrantz broadcasting Guildenstern's disposition to others so that Guildenstern is socially ostracized even further in the future. If Rosencrantz made a mistake with respect to Guildenstern's disposition then Rosencrantz would suffer from being exploited by Guildenstern. One question to bear in mind, for it will prove to be important later, is what would Rosencrantz's response be if she did not know what disposition Guildenstern adopted?

Now consider an encounter between Rosencrantz and Becassine,

---

self-interest and Gauthier's conception of justice may test our considered intuitions.

[44] Morals By Agreement; op cit., p. 174.

where Becassine knows by surreptitious means (and Rosencrantz doesn't) that Rosencrantz will be deported back to his own country before she, Becassine, can complete her part of the agreement. Once again assume that translucency enables each to detect the other's disposition; [45] both know that the other has a disposition to cooperate and therefore they agree to the cooperative agreement. Rosencrantz carries out her part of the agreement and Becassine only goes through the motions of carrying out her part of the agreement. Once Rosencrantz has been deported back to the old country Becassine ceases her cooperative activities and reaps the reward of having had Rosencrantz assist her without paying the cost of mutual reciprocation. In this case, by taking a "moral holiday"[46] when it was convenient to do so, Becassine's expected utility would be greater than Rosencrantz's.

Now suppose the situation was completely reversed in such a way that Rosencrantz could get away with not carrying out her part of the cooperative arrangement for Becassine was about to be deported. The question we need to ask is why wouldn't Rosencrantz defect from this agreement when she could do so without suffering

---

[45] One might want to raise the objection that although translucency can detect another's disposition to cooperate that it may not be able to detect a C-S's latent tendency for defection. This issue will be addressed in chapter III but for the moment I assume that translucency cannot detect this latent tendency for defection.

[46] Hare, R.M.; Moral Thinking:Its Levels, Method and Point; Clarendon Press, Oxford; 1981; p. 57.

any future loss? The answer, I suggest is that Rosencrantz is, by Gauthier's account, morally incontinent; she acts morally when she need not do so. Moreover, by suffering from this moral incontinence, Rosencrantz's expected utility must be less than Becassine's expected utility.

The above example highlights the intuition on which I began to construct my argument against Gauthier. That is to say I wanted to build an argument to show that if one could defect from one's agreements without suffering any costs then a rational agent would do so. In order to make such an argument I need to spell out in detail the differences between these dispositions and then I need to find a formal argument which will support this intuition. I attempt both of these tasks in the following chapters.

Chapter II

The Motivational Component

Strategic rationality, as opposed to parametric rationality, involves interaction with another person where not only is each person in the strategic interaction concerned to maximize his or her utility but also each person's choice is reciprocally anticipated by each party to the interaction.[1] Consider the following example:

> Jane wants very much to go to Ann's party. But even more she wants to avoid Brian who may be there. Brian wants very much to avoid Ann's party. But even more he wants to meet Jane. If Jane expects Brain to be at Ann's party she will stay at home. If Brian expects Jane to stay at home then so will he. If Jane expects Brain to stay at home she will go to the party. If Brian expects Jane to go so will he. If Jane ... but this is where we began.[2]

Jane and Brian have a problem, the same problem we all face when we are each concerned to act in our own best interest and

---

[1] Recall conditions A, B and C from the Introduction:

A: Each person's choice must be a rational response to the choices she expects the others to make.

B: Each person must expect every other person's choices to satisfy condition A.

C: Each person must believe her choice and expectations to be reflected in the expectations of every other person.

[2] Morals By Agreement; op cit., p. 60.

where the outcome which best manifests that interest is directly dependant on what someone else does. Gauthier must resolve this problem of strategic interaction and he must do so within the confines of the idealizing conditions A, B and C.

Some might assume that Jane's and Brian's problem is supposed to mirror the problem found in Prisoner's Dilemma (PD) type situations. However this is not the case. Gauthier uses this Jane and Brian example as an initial step in moving towards a solution of the Prisoner's Dilemma type situation. He does not, with this example, resolve the Prisoner's Dilemma but rather points towards an approach which might resolve it,[3] i.e., by defining "rational response" in Condition A as "optimising response" in place of "utility maximizing response."[4]

---

[3] Gauthier thinks he resolves the Prisoner's Dilemma with his argument for constrained maximization but in these remaining two chapters I argue that he is unsuccessful in his attempt, but this is to get ahead of myself. Gauthier's formal apparatus for dealing with the Jane and Brian problem is explicated in Appendix A.

[4] Note that in condition A "rational response" is undefined. Granted one would think that "rational response" would mean, given Gauthier's definition of rationality, "utility maximizing response," and this is how Gauthier initially defines "rational response" in this first condition. However, he later redefines rational response to mean an "optimizing response" and this latter condition is referred to as A'. The purpose of the redefinition of "rational response" is to circumvent a problem which arises with the first definition. If Gauthier left "rational response" defined simply as a "utility maximizing response" his argument would suffer from the obvious Pareto-Optimality objection, i.e., not all utility maximizing responses are optimal responses. By making the change to A' from A, i.e., by redefining "rational response" to mean "optimizing response," Gauthier can claim that while it may be the case that not all utility maximizing responses are optimal, it is the case that all

Recall that the two prisoners in the Prisoner's Dilemma are separated and unable to communicate with each other. In terms of years in jail they face the following situation. If both prisoners confess then they each face 5 years in jail. If Ed confesses and Fred does not then Ed faces 1 year in jail and Fred faces 10 years in jail. If Fred confesses and Ed does not then Fred faces 1 year in jail and Ed faces 10 years in jail. If neither of them confess, which is the best outcome, then they each get 2 years.

One can easily illustrate that Gauthier's Jane and Brian example does not mirror the Prisoner's Dilemma by presenting a matrix for each which outlines the preferences of each party to the strategic interaction.

Prisoner's Dilemma

	C	NC
C	5,5	1,10
NC	10,1	2,2

Jane & Brian

	G	S
G	4,1	1,4
S	2,3	3,2

In the Prisoner's Dilemma matrix we have a cell that is mutually preferred by both prisoners, i.e., the cell containing 2,2. In the Jane and Brian example we have no such cell which is

---

optimizing responses are utility maximizing responses, at least they are in the long run. Then, the argument continues, one can better maximize one's utility, overall, by adopting an optimizing response over a utility maximizing response.

mutually preferred. There is another distinction as well between these two matrixes; in the Prisoner's case there is a dominant solution, i.e., a solution that either prisoner would opt for regardless of what the other does, i.e., the cell where they both confess. In the Jane and Brian example we have no such dominant solution.

In analyzing the example of Jane and Brian we find that each participant, Jane and Brian, have four preferences. Jane's and Brian's preferences, hierarchically ordered, are as follows:

J1. Jane goes to the party and Brian stays at home.
J2. Jane stays at home and Brian goes to the party.
J3. Jane and Brian both stay at home.
J4. Jane and Brian both go to the party.

B1. Brian and Jane both go to the party.
B2. Brian and Jane both stay at home.
B3. Brian goes to the party and Jane stays home.
B4. Brian stays at home and Jane goes to the party.

The problem, it will be remembered, is for Gauthier to formulate a principle of strategic rationality that satisfies the three conditions, A, B, and C, of strategically rational choice.[5] Each of the participants has two choices; they can either go to

---

[5] An important point to note, regarding what has been said thus far concerning conditions A, B, and C, is that Gauthier has defined "rational response" in condition A as "utility maximizing." By defining rational response in condition A as "utility maximizing" Gauthier builds on his account of practical rationality and he can more easily illustrate the central problem of strategically rational choice. He will change this definition as his argument proceeds, as I have already pointed out; rational response will come to mean an "optimizing response."

the party or they can stay home. Consider the following as concerns Jane:

If J1 then, via condition A, Jane expects Brian to stay at home. But by condition B she must expect Brian to expect her to stay at home. This, however, violates condition C because her choice is not reflected in what she believes to be Brian's expectations.

If J2 then, via condition A, she must expect Brian to go to the party. But by condition B she must expect Brian to expect her to go to the party. Once again this violates condition C on the grounds that Jane's choice is not reflected in what she believes to be Brian's expectations.

The same argument holds for Brian; he can either go to the party or he can stay at home. If B1 then, via condition A, Brian expects Jane to go to the party. But by condition B he must expect Jane to expect him to go to the party. But if Jane expects Brian to go to the party then condition C is violated because his choice is not reflected in what he believes to be Jane's expectations.

If B4 then, via condition A, he must expect Jane to go to the party. But by condition B he must expect Jane to expect him to go to the party. As with Jane condition C is violated because

Brian's choice is not reflected in what he believes to be Jane's expectations.

In cases of strategic interaction each agent's behaviour is reciprocally anticipated by the other. This mutually reciprocal anticipation is manifested in both the Jane and Brian example and the two-person PD and it gives us grounds for raising the spectre of the compliance problem. How can each of the participants in these two examples ensure the optimal outcome given that their expected behaviour will be anticipated and responded to by the other?

We have seen that there are two tasks which Gauthier must accomplish with respect to the motivational component of his theory. First he must solve the compliance problem; he must show that it is rational, i.e., utility maximizing, for rational agents to comply with the agreements they have made. Furthermore, he must show that his answer to the compliance problem, constrained maximization, is also the answer to the question of which strategy is the uniquely rational strategy, which is to say that no other strategy will generate as much overall utility as does constrained maximization. Gauthier believes constrained maximization will accomplish both of these tasks given translucency and narrow compliance.[6]

---

[6] I will not address the issue of translucency in any great detail in this chapter or in the next chapter for any objections to translucency have to be real-world objections and therefore

Gauthier's approach for solving the compliance problem and the problem of which disposition is the uniquely rational strategy is a threefold approach: first he reinterprets the received view (i.e., Bayesian decision theory and the Von Neumann-Morgenstern theory of games)[7] of what a utility-maximizing conception of rationality is, he then makes a distinction between constrained maximizers (CM's) and straightforward maximizers (SM's) of utility and finally he strengthens his position with his arguments for translucency and narrow compliance.

I will not argue against the overall approach Gauthier utilizes. On the contrary, I will make use of his approach and argue, using some of Gauthier's own arguments, first that constrained maximization and straightforward maximization, CM and SM, are not jointly exhaustive dispositions. I will argue that a third disposition, call it a C-S disposition, can be identified and distinguished from both of Gauthier's constrained and straightforward maximizing dispositions. Moreover, I will argue

---

they would not have much force against Gauthier's idealized argument. What I do have to say against translucency is simply to point out problems that might exist with the concept conceding that Gauthier could quite easily respond to the problems that I will identify. I will not be addressing the issue of narrow compliance for it is not crucial, at this stage of the argument, to either Gauthier's or my position. For a detailed criticism of narrow compliance see Kraus and Coleman, op cit.

[7] See Appendix A for a detailed account of what is involved in Gauthier's answer.

that not only is this third disposition at least as rational as Gauthier's CM disposition, in some cases, but also it reflects a more plausible conception of what one would expect of rational agents who are intent on maximizing their utility.

Gauthier leaves himself open to criticisms by counter-example by arguing that his CM disposition is the uniquely rational strategy. I will not argue that this third C-S disposition is the uniquely rational strategy, but rather I will argue for the weaker claim that it is at least as rational as Gauthier's CM disposition in some cases. That is to say I can find conditions under which it would be more rational, in terms of expected utility maximization, to choose a C-S disposition rather than a CM disposition. I waive presenting this argument until the next chapter when I come to consider Gauthier's argument that constrained maximization is the uniquely rational strategy.

In this chapter my arguments will focus on arguing for this C-S disposition in terms of distinguishing it from Gauthier's two alternatives. I will argue that there is both a behavioral difference and a dispositional difference which one can identify, and these are sufficient to make the distinction. This does not necessarily imply that the behaviors of the three alternative dispositions will always be different for they will not be. On the other hand there will always be a dispositional difference

which is readily identifiable by dissecting their respective psychologies.

The received view of a utility-maximizing conception of rationality identifies rationality with utility-maximization at the level of particular choices such that "a choice is rational if and only if it maximizes the actor's expected utility." Gauthier, on the other hand, identifies rationality with utility-maximization at the level of dispositions to choose such that "a disposition is rational if and only if an actor holding it can expect his choices to yield no less utility than the choices he would make were he to hold any alternative disposition."[8] This point is essential to Gauthier's argument, for he will go on to argue that the disposition one chooses affects the situations one can expect to encounter. A person who chooses a CM disposition can expect to be welcomed into a community of similarly disposed CM's while a person who chooses an SM disposition will suffer from the consequences of social ostracism.

In order to make my case for a C-S disposition being a counter-example to Gauthier's CM disposition I must accept Gauthier's reinterpretation of what a utility maximizing conception of rationality comprises. Thus I too identify rationality at the level of dispositions to choose and I must show that the choices which a person with a C-S disposition would

---

[8] Morals By Agreement; op cit., pp. 182-183.

make will yield no less utility than the choices he would make if he were to hold a CM disposition.

Moreover, in what follows I will be compelled to accept, on behalf of my C-S disposition, many things Gauthier says about CM's. My argument for this C-S disposition will lie in arguing that although a C-S disposition is similar to both a CM and an SM disposition in some respects, it is nonetheless a distinct disposition. One will be able to distinguish a C-S disposition from either a CM or an SM disposition in at least one of two ways: by their behaviour in a given situation or by their reasoning in that situation. One might think that they could also be distinguished by the reasons for their particular behaviour but unfortunately they cannot. If I were to try and make the distinction on the grounds of their reasons for their particular behaviour, this would involve a denial of Gauthier's definition of rationality as utility maximization. Their reasons for acting are the same: each tries to maximize his own utility.

Constrained maximization is the motivational component of Gauthier's theory and therefore his answer to the market compliance problem of why Q would comply with the agreement made with P once he, Q, is out of the water and on terra firma, of why Sam McGee would comply with his agreement with Grasp the Banker, and the formal problem illustrated by the Prisoner's Dilemma. Constrained maximization is also Gauthier's answer to Hobbes's

Foole.

The first step in Gauthier's detailed argument to solve the problem of the motivational component of his theory is to make a distinction between an "individual strategy" and a "joint strategy." An individual strategy, Gauthier says, is "a lottery over the possible actions of a single actor," and a joint strategy is "a lottery over possible outcomes." [9] Cooperation, then, can be understood as agreeing to a joint strategy as opposed to an individual strategy.

"This is fine," says Hobbes's Foole (and Gauthier's rational agents), "but when is it rational to cooperate? I think it is rational to cooperate if and only if the utility I expect from

---

[9] Morals By Agreement; op cit., p. 166; also see Appendix A. A "strategy" is defined as "a lottery over possible actions." There are two types of strategies; the first type of strategy is a **pure strategy**. This is where one assigns the probability of 1 to one action and 0 to all other actions. This type of strategy has only one prize and the prize is always awarded. The second type of strategy is a **mixed strategy**; it is a lottery with several actions as prizes. In a mixed strategy one assigns a non-zero probability to more than one action with the sum of probabilities assigned being equal to 1.

Gauthier makes two innocent idealizing assumptions with respect to lotteries. They are innocent assumptions as long as they are confined to ideal theory; if they were ported over into partial compliance theory then they would lose their status of being innocent assumptions. The two assumptions are that there is "universal availability of a randomizing device" and that this device is "costless to use." Note that Gauthier, by expanding the scope of choice from actions to strategies, has not increased the range of possible outcomes. This is the case because strategies are simply lotteries over possible actions and, regardless of the outcome of the lottery, there is still only one possible action.

cooperation, i.e., from adopting a joint strategy as opposed to an individual strategy, is at least equal to what I might expect were I to act on my best individual strategy." "Therefore," the Foole continues, "this defeats the end of cooperation." "Cooperation," he adds, "presupposes that the expected utilities are to everyone's benefit." "But," the Foole goes on, "why should I be concerned whether or not everyone benefits?" "If my expected utility, from adopting an individual strategy, is at least equal to my expected utility from adopting a joint strategy, and if I am not concerned whether everyone benefits or not, then why is it rational for me to choose a joint strategy as opposed to an individual strategy?"

The problem with the Foole's argument, Gauthier argues, is that the Foole misses the point. His individual strategy is rational if, and only if, it maximizes his utilities given the strategies adopted by the others. The Foole's utilities are dependent on what strategies the other person adopts; they may be more, less, or nonexistent, depending on which strategy the other person adopts. A joint strategy, however, is not like this. "A joint strategy is rational only if (but not if and only if) it maximizes one's utilities given the utilities afforded to the other persons." [10]

This brings us to the second, and the most important step,

---

[10] Morals By Agreement; op cit., p. 167.

in Gauthier's argument, the distinction he makes between constrained maximizers and straightforward maximizers of utility.

> Let us say that a straightforward maximizer is a person who seeks to maximize his utility given the strategies of those with whom he interacts. A constrained maximizer, on the other hand, is a person who seeks in some situations to maximize her utility, given not the strategies but the utilities of those with whom she interacts. The Foole accepts the rationality of straightforward maximization. We, in defending condition A' for strategic rationality, accept the rationality of constrained maximization. [11]

I argue that C-S's, like CM's, will also seek to maximize their utilities given the utilities of others.

A CM is not simply an agent who seeks to maximize her utilities given the utilities of those with whom she interacts; a CM has some other characteristics as well. Gauthier offers this enhanced view of the CM: A CM is

> (i) someone who is conditionally disposed to base her actions on a joint strategy or practice should the utility she expects were everyone so to base his action be no less than what she would expect were everyone to employ individual strategies, and approach what she could expect from the cooperative outcome determined by minimax relative concession; (ii) someone who actually acts on this conditional disposition should her expected utility be greater than what she would expect

---

[11] Morals By Agreement; op cit., p. 167. "Condition A'" was discussed earlier. The reader will recall that Condition A of strategic rationality stated "each person's choice must be a rational response to the choices she expects the others to make," where rational response was understood as utility maximizing response. Condition A' has optimizing response substituted for utility maximizing response. Also see Appendix A.

were everyone to employ individual strategies.[12]

Basically a CM is ready to cooperate if she expects others to cooperate and the outcome is beneficial and not unfair; moreover, she will cooperate if she expects to benefit. In order to determine that she will, in fact, benefit she must take into account that some persons, SM's, may not cooperate.

A C-S is somewhat like the CM with respect to both of these two characteristics. He is conditionally disposed to base his actions on a perceived joint strategy or practice, as opposed to an actual joint strategy,[13] should the utility he expects be no less than what he would expect were everyone to employ individual strategies, and approach what he could expect from the cooperative outcome determined by minimax relative concession; and he is someone who actually acts on this conditional disposition to base his actions on a perceived joint strategy or practice should his expected utility be greater than what he would expect were everyone to employ individual strategies. Thus a CM disposition and a C-S disposition are quite similar so far.

---

[12] Morals By Agreement; op cit., p. 167. This second condition of constrained maximizers highlights the epistemological problem I identified in Chapter 1. Given that I adopt Gauthier's second condition on behalf of my agent with the C-S disposition I also face this problem, but, for the sake of argument, I waive trying to defend against this objection.

[13] The reason for the distinction between the perceived versus actual joint strategy will become clearer as I proceed. It should be simply noted for the time being that the CM perceives that the C-S has adopted an actual joint strategy.

There are three additional points Gauthier wants to establish about CM's that will add to his overall argument. These three points will also pertain to my view of a C-S disposition and they will also add to my overall argument. The first is that the range of acceptable joint strategies not be specified. Not all CM's will cooperate in every acceptable joint strategy; some might not cooperate in a given joint strategy for the reason that if they don't cooperate in this particular joint strategy they might obtain "agreement on, or acquiescence in, another joint strategy which in being fairer is also more favorable to oneself."[14] Furthermore, there may very well be cases of joint strategies that are not completely acceptable, yet the CM may choose to act cooperatively if the case in point affords her an expected utility "approaching what she would expect from fully rational cooperation." [15]

Gauthier argues that if the range of acceptable joint strategies is not completely specified beforehand in ideal theory, then we can more readily tie the conclusions drawn in ideal theory to the real world of ships, sealing wax, cabbages and kings. This argument, though important for relating his ideal

---

[14] Morals By Agreement; op cit., p. 168.

[15] Morals By Agreement; op cit., p. 168. To say that cooperation is "fully rational" is to say that it leads to a Pareto-optimal outcome, i.e., there is no possible outcome affording some person a greater utility and no person a lessor utility.

theory to partial compliance theory, does not play a significant role in what we are considering at the moment.

The second point relates to the number of persons expected to cooperate before a CM will base her actions on a joint strategy (and a C-S will base his actions on a perceived joint strategy). It is not the case that they will base their actions on a joint strategy (or a perceived joint strategy) whenever a nearly fair and optimal outcome would result were everyone to do likewise; rather "her disposition, (i.e., the CM), to cooperate is conditional on her expectation that she will benefit in comparison with the utility she could expect were no one to cooperate," [16] i.e., universal non-cooperation. This is to say that the CM and the C-S both estimate "the degree to which others will cooperate" before they make up their mind. [17]

If a CM bases her own cooperative behaviour on her estimation of other's cooperative behaviour then so does a C-S. By basing their cooperative behaviors on their estimation of others' cooperative behaviours then both the CM and C-S protect themselves from being exploited by SM's. However, if one has adopted a constrained maximizing disposition she may find herself in situations such that she will be made worse off than if she had never entered into the cooperative arrangement. Nevertheless,

---

[16] Morals By Agreement; op cit., p. 169.

[17] Morals By Agreement; op cit., p. 169.

a CM would still cooperate, Gauthier argues, for "given her ex ante beliefs about the dispositions of her fellows and the prospects of benefit, participation in the activity affords her greater expected utility than non-participation."[18] I argue that a person with a C-S disposition would also still cooperate for the same reasons.

It may seem strange for Gauthier and myself to argue that a CM and a C-S would still cooperate even if, possibly due to some unforeseen circumstances, they became worse off than if they had never entered the arrangement. However, one must remember two things. First, CM's and C-S's have adopted dispositions to comply; they have not chosen a simple act-compliance strategy. Second, a CM and a C-S base their choice, not on the strategies but rather the utilities of those with whom they interact. Thus both would still carry through with a cooperative arrangement,[19] where they stand to be made worse off, because they have adopted a disposition and because their expected utility from cooperation is greater than their expected utility from universal non-cooperation.

This argument of Gauthier's is a testimony to what it means to choose a disposition. For Gauthier, choosing a disposition

---

[18] Morals By Agreement; op cit., p. 169.

[19] This claim will be qualified with respect to a C-S disposition as I proceed.

means being conditionally disposed to act on that disposition whenever certain conditions obtain. For a CM this means, basically, if he believes himself to be amongst other similarly disposed persons he will base his actions on a joint strategy, i.e., act cooperatively. Choosing a disposition is a case of parametric choice. A rational agent reasons parametrically, when considering which disposition he will have, to his own best disposition. [20] Once having made his choice of dispositions he will act on that disposition as long as the conditions for acting on it obtain. Of course, this is not to say that he will act on every occasion that these conditions obtain, but rather he will act on those occasions which maximize his expected utility in terms of his coherent and considered preferences. This is to say that once a person chooses a disposition he accepts certain necessary and sufficient conditions for acting on that disposition; he will act on that disposition given that the necessary and sufficient conditions obtain and his expected utility coheres with his coherent and considered preferences. All actors, regardless of which disposition they have chosen, accept coherence with one's preferences as a necessary condition for acting; this is a formal condition governing ideal actors engaged in strategic interaction.

The third point Gauthier wants to make about CM's is simply

---

[20] Morals By Agreement; op cit., pp. 170-171.

that CM's are not merely SM's in their "most effective disguise."[21] The CM really has adopted a disposition to comply in those cases where her expected utility from cooperation is greater than her expected utility from universal non-cooperation. It's not the case, Gauthier argues, that a CM is simply an SM who has taken a longer and closer look at the future such that she would sacrifice some immediate benefits for the greater expected benefits generated from defecting later on, after having established her trustworthiness within a community of similarly disposed CM's. Gauthier's argument for this conclusion is that

> The constrained maximizer does not reason more effectively about how to maximize her utility, but reasons in a different way. We may see this most clearly by considering how each faces the decision whether to base her action on a joint strategy. The constrained maximizer considers (i) whether the outcome, should everyone do so, be nearly fair and optimal, and, (ii) whether the outcome she realistically expects should she do so affords her greater utility than universal non-cooperation. If both of these conditions are satisfied she bases her actions on the joint strategy. The straightforward maximizer considers simply whether the outcomes he realistically expects should he base his actions on the joint strategy affords him greater utility than the outcome he would expect were he to act on any alternative strategy -- taking into account, of course, long-term as well as short-term effects. Only if this condition is satisfied does he base his action on the joint strategy.[22]

---

21 Morals By Agreement; op cit., p. 169. I will argue, after I have described more fully what is involved in having a C-S disposition, that a person with the C-S disposition is also not an SM in her most effective disguise even though some might think she is.

22 Morals By Agreement; op cit., p. 170.

When Gauthier argues that a CM does not reason more effectively than an SM but rather reasons in a different way I say the same with respect to a C-S. However, I go further. A C-S doesn't reason more effectively than a CM but rather reasons in a different way than a CM. Conditions (i) and (ii) are necessary and sufficient conditions for a person with a CM disposition to base her actions on a joint strategy. If both of these conditions obtain then a person with a CM disposition will act cooperatively.

A person with a C-S disposition takes into account both conditions (i) and (ii) described above, as does the CM, but these two conditions are, at times, merely necessary conditions for a person with a C-S disposition; at other times these two conditions will prove to be both necessary and sufficient. Let us first look at those occasions where these two conditions are merely necessary conditions before we look at those occasions where these two conditions are both necessary and sufficient.

In order for conditions (i) and (ii) to be simply necessary conditions for a person with a C-S disposition I need to introduce a third condition, condition (iii), which has to do with whether or not a defection at the opportune moment will adversely affect his future opportunities for cooperative interaction. If a defection, i.e., acting straightforwardly at

the opportune moment, will adversely affect his opportunities for being a party to some future agreement then he will not defect; if a defection will not adversely affect his opportunities for being a party to some future agreement then he will defect. When defecting at the opportune moment does not adversely affect his opportunities for future cooperative interaction conditions (i) and (ii) will prove to be only necessary conditions for a person with a C-S disposition to base his actions on a perceived joint strategy. When defecting would adversely affect his chances for future cooperative interaction conditions (i) and (ii) will prove to be both necessary and sufficient conditions for a person with a C-S disposition to base his actions on a joint strategy.

Note that when I say defecting at the opportune moment will adversely affect his opportunities for future cooperative interaction conditions (i) and (ii) will be only necessary conditions for a person with a C-S disposition to base his actions on a perceived joint strategy. It is not the case that he bases his actions on an actual joint strategy, for an actual joint strategy is defined as adopting a course of action over time such that, for example, the agent is expected to do all actions from A_1 up to and including A_n. Anything which falls short of A_n cannot be a true joint strategy, thus what the C-S does, i.e., the C-S who finds that defecting will not adversely affect his future opportunities for cooperative interaction, is act such that he carries out all the actions required of the

joint strategy except A_n. He is perceived, by the CM, as acting on a joint strategy whereas he is really acting on an individual strategy. Note, however, by having the person with the C-S disposition acting in this manner, I can still distinguish him from an SM "in his most effective disguise." [23]

A person with a C-S disposition, like a person with a CM disposition, always reasons such that he takes into account conditions (i) and (ii) whereas a person with an SM disposition never considers either of these two conditions. The person with an SM condition never concerns himself with whether the outcome is nearly fair and optimal nor whether the outcome he realistically expects affords him greater utility than universal non-cooperation. The person with the SM disposition only considers whether the outcome he realistically expects is greater than the outcome from any alternative strategy.

When condition (iii) is met I stipulate that one's opportunities for future cooperative interaction are not adversely affected, but this requires some further elaboration. It may be the case that, even though one's particular defection from the perceived joint strategy is not traceable back to the actual defector, it will be known that a defection has occurred. This raises a potential problem for my argument in terms of the number of defections which are possible. Am I to consider the

---

[23] Morals By Agreement; op cit., p. 169.

number of possible defections on the part of the person with the C-S disposition to be great in number or only very rare?

The problem arises if I consider the number of defections on the part of the person with the C-S disposition to be great in number. If the opportunities for defection on the part of a person with a C-S disposition were sufficiently great then it might offset the benefits of engaging in cooperative interaction in the first place. Everyone would be adopting perceived joint strategies and then acting straightforwardly when the opportune moment arose. If this were the case one might be better off to adopt an SM disposition, rather than a CM disposition, (and perhaps a C-S disposition if Gauthier would have considered this third alternative disposition) in order to mitigate the losses suffered. However, if everyone did adopt an SM disposition then the compliance problem would be unresolved and a political solution, some substantive rule of action, would be called for. A sovereign, perhaps Hobbes' Leviathan, would, with "well placed kicks," [24] coordinate and control everyone's behaviour thus mitigating the damage suffered from everyone adopting an SM disposition.

Suppose I consider the opportunities for defection on the part of those with a C-S disposition to be rare in number, then I can argue that a person with a C-S disposition can expect to do

---

[24] Morals By Agreement; op cit., p. 163.

at least as well in some cases as a person with a CM disposition, for, as I will argue, a person with a C-S disposition has all the opportunities for cooperative interaction that a person with a CM disposition has, and thus all the benefits, but also the person with a C-S disposition has the additional benefits of defecting whenever defecting at the opportune moment will not adversely affect his opportunities for future cooperative interaction.

Gauthier identified a need for morality with his argument for the morally free zone showing that the world was not the perfectly competitive market that the laissez-faire capitalists assumed it to be. There were, he argued, economic inefficiencies, free riders and parasites. Recall also that in a perfectly competitive market, Pareto-optimality is guaranteed: what one person does, in a perfectly competitive market, in no way affects the optimality of the outcome for another person. Given that there are these economic inefficiencies and Pareto inefficient outcomes, Gauthier was able to identify a need for morality. My argument is simply that by identifying a need for morality Gauthier opened himself up to the possibility of defections occurring, specifically the defections endorsed by the person with the C-S disposition.

As far as the actual number of such possibilities is concerned it need not be specified for the reason that the possibility alone is sufficient to make my argument. One must

remember that a rational agent is reasoning parametrically about which disposition he will adopt. Thus he considers a world where there is the possibility of defecting at the opportune moment such that his defection will not adversely affect his opportunities for future cooperative interaction. He considers a world where this condition might be met for himself, and not for the other person, and he calculates his expected utility given that assumption. He also considers the possibility of a defection occurring where neither party's opportunities for future cooperative interaction are not jeopardized and he calculates his expected utility given this expectation. But, if the possibility of defecting at the opportune moment, where such a defection does not adversely affect one's opportunities for future cooperative interaction, is rare for one person then the possibility of this condition being met for both persons in the cooperative venture is rarer still, and the rational agent is cognizant of this when he calculates his expected utilities.

When defecting at the opportune moment will not adversely affect his opportunities for future cooperative interaction conditions (i) and (ii) are merely necessary conditions for a person with a C-S disposition to base his actions on a perceived joint strategy and condition (iii) is the sufficient condition. When defecting will adversely affect his opportunities for future cooperative interaction conditions (i) and (ii) are both necessary and sufficient conditions for a person with a C-S

disposition to base his actions on an actual joint strategy. This is to say that if defecting will adversely affect his opportunities for future cooperative interaction a person with a C-S disposition will base his actions on an actual joint strategy on every occasion that a person with a CM disposition would; there would be no behavioral difference between the two actors but there would be a dispositional difference, for the person with the C-S disposition would determine if defecting would, or would not, adversely affect his opportunities for future cooperative interaction. If his future opportunities would not be adversely affected a person with a C-S disposition would base his actions on a perceived joint strategy in every case that a person with a CM disposition would base her actions on an actual joint strategy. But in this case there would be a behavioral difference between the two, for the person with the C-S disposition would defect at that time when his defection would not be discovered, thereby ensuring that his opportunities for future cooperative interaction would not be adversely affected.

In the case of a CM the behaviour would involve basing one's actions on a joint strategy and carrying out the terms of the particular agreement, A_1 up to and including A_n. In the case of the C-S, when defecting at the opportune moment would not adversely affect his opportunities for future cooperative interaction, the behaviour would involve basing his actions on a perceived joint strategy, A_1 up to but not including A_n, thus

ensuring the cooperative benefits, and then defecting from the perceived joint strategy at the opportune moment when the defection would not be discovered.

I say that a person with a C-S disposition would base his actions on a perceived joint strategy as opposed to an actual joint strategy, when defecting at the opportune moment would not adversely affect his opportunities for future cooperative interaction, because if he did not there would be no cooperative benefits to be realized by his defection, and moreover his disposition would simply be an SM disposition. Furthermore, if he didn't base his actions on the perceived joint strategy but rather based his actions on an individual strategy he would suffer from the same social ostracism that Gauthier says the person with the SM disposition would, and consequently his future opportunities for cooperative interaction would be affected.

The C-S is willing, like the CM, to base his action on an actual joint strategy if only conditions (i) and (ii) obtain; these are necessary conditions for both of them but they are also sufficient conditions for the CM. If condition (iii) is met such that defecting at the opportune moment will not adversely affect his opportunities for future cooperative interaction the person with the C-S disposition has the best of both worlds; he bases his actions on a perceived joint strategy and defects at the appropriate moment, secure in the knowledge that his defection

will not adversely affect his future opportunities for cooperative interaction. If condition (iii) is not met, defecting will adversely affect his opportunities for future cooperative interaction, then conditions (i) and (ii) are both necessary and sufficient for him and he remains willing to base his actions on an actual joint strategy and is thus indistinguishable behaviorally from a CM. A CM, who does not take into account condition (iii), would not act straightforwardly as long as conditions (i) and (ii) obtain.

If condition (iii) is not met, (defecting will adversely affect his opportunities for future cooperative interaction), there will still be a dispositional difference between the CM and the C-S. For the former takes conditions (i) and (ii) as necessary and sufficient, while the latter only takes them as necessary conditions (when condition (iii) is met) and necessary and sufficient when condition (iii) is not met. Moreover, when condition (iii) is not met there is a behavioral difference between the two. The CM bases her actions on a joint strategy, i.e., acts cooperatively, and reaps the rewards afforded her for doing so. The C-S, on the other hand, bases her actions on a perceived joint strategy, i.e., she too acts cooperatively, but when the opportune moment arrives the person with the C-S disposition defects from the perceived joint strategy.

She has to defect from the perceived joint strategy, as

opposed to simply acting straightforwardly from the start, for two reasons. First, as I have already hinted at, if she simply acted straight-forwardly from the start she would be indistinguishable from an SM; this would have the result of discounting my claim that a CM and an SM disposition are not jointly exhaustive dispositions. Second, in order to reap the rewards of her defection, her defection must be from a perceived joint strategy. That is to say, by basing her actions on the perceived joint strategy certain benefits come into being by virtue of the cooperative enterprise; these benefits only arise because of the cooperative enterprise and her defection from the perceived joint strategy allows her to reap the rewards of these benefits. If she simply acted straightforwardly, based her action on an individual strategy like the SM, the opportunity to obtain these cooperative benefits through defection would never arise.

One may object at this point that the potential defection on the part of my C-S is circumvented by Gauthier's assumption of translucency. CM's would cultivate their ability to detect another's disposition to cooperate, as Gauthier says they would,[25] and therefore even if a C-S took into account condition (iii) it would be to no avail. CM's would have nothing to do with him for they would have detected this latent tendency to defect in the same way that they would detect an SM disposition. This objection raises an important point, for it emphasizes the fact that there

---

[25] Morals By Agreement; op cit., pp. 180-181.

is more riding on translucency than Gauthier has admitted and perhaps more than he even realized. However, having said this, one needs to point out that any objection to translucency is a real-world objection and not an objection to Gauthier's idealized argument where the agents are fully rational and fully informed. But, having said this, it still can be said that real world objections can serve pedagogical purposes and therefore the objection should be examined while bearing in mind that Gauthier could respond to these real-world objections with a more accurate and detailed description of what is involved in the concept of translucency.

Gauthier says translucency detects another's "disposition to cooperate," [26] but the important question is, does translucency also detect the latent tendency for defection on the part of the person with the C-S disposition? The person with the C-S disposition has a disposition to cooperate and this disposition to cooperate is manifested whenever conditions (i) and (ii) are both necessary and sufficient conditions for him to base his actions on a joint strategy, which occurs most of the time. The person with the C-S disposition still has a disposition to cooperate even when defecting from a perceived joint strategy will not adversely affect his opportunities for future cooperative interaction; he doesn't lose his disposition for cooperation in those rare cases when he defects as the CM does

---

[26] Morals By Agreement; op cit., p. 174.

not lose his disposition to cooperate whenever he finds himself amongst SM's and acts straightforwardly. Therefore, given the distinction I have drawn between the CM and the C-S disposition, translucency, as Gauthier has defined it, should not be able to identify this latent tendency for defection on the part of the person with the C-S disposition. Some might want to argue that by making this latent tendency for defection undetectable that I have made my persons with the C-S disposition to be opaque and therefore, given the ideal nature of Gauthier's argument, the competing dispositions are not on an equal footing. One way of circumventing this particular objection is to concede it for the moment and look at a case where the agents are neither translucent nor opaque but rather completely transparent.

Assume then a case where completely transparent CM's, SM's and C-S's are in a PD type situation and recall the relevant structure of such dilemmas; each person in such a two-person dilemma benefits from mutual cooperation in relation to mutual non-cooperation, but each person benefits from non-cooperation whatever the other does. Gauthier's argument is that CM's would cooperate given that conditions (i) and (ii) are met; SM's would not cooperate for they would each expect the other to base his actions on an individual strategy.[27] My argument is not as

---

[27] Note that this argument, and the ones to follow, completely ignores the three conditions, A, B, and C, of strategically rational choice. For Gauthier's solution to the formal problem of strategically rational choice see Appendix A.

simple: if C-S's have taken condition (iii) into account and both have found that defecting will adversely affect their future opportunities for cooperative interactions then they will cooperate. If both have found that defecting will not adversely affect their opportunities for future cooperative interaction then they will not cooperate. Given the transparency assumption either C-S knows whether or not the other will defect; if either plans on defecting then they will not cooperate and hence end up with the same outcome as SM's. In the first case two C-S's end up with the same outcome as two CM's while in the latter two cases two C-S's end up with the same outcome as two SM's.[28]

By assuming transparency, as opposed to translucency, I have shown that in some cases choosing a C-S disposition would be just as rational as choosing a CM disposition for in those cases C-S's would have an expected utility equal to that of CM's. However, by assuming transparency rather than translucency I have denied the possibility of C-S's ever obtaining the extra utility expectation generated from an individual defection when defecting will not adversely affect one C-S's opportunities for future cooperative

---

[28] By assuming transparency we have ignored a potential problem. What happens, if we assume translucency, and some party, (or both parties), does not know what disposition the other has? After all, as Gauthier says, translucency detects another's disposition "not with certainty, but as more than mere guesswork." This leaves open the following two possibilities, first, one might be mistaken, and second, one might not know which disposition the other has chosen. This problem will be addressed in the following chapter when I come to consider Gauthier's translucency argument.

interaction.

Let us now assume translucency and consider first the case where this latent tendency for defection is undetectable and then the case where this latent tendency for defection is detectable, and consider whether a C-S disposition is still at least as rational a strategy in some cases as a CM disposition. When I say 'let us assume that this latent tendency for defection is undetectable,' I mean only that. One can still detect, via the virtue of translucency, whether or not another is an SM or a CM. What one cannot detect is whether or not a C-S is actually a CM or actually a C-S.

Gauthier's argument for CM's is the same: CM's will enjoy opportunities for cooperative interaction which SM's will not, because CM's will detect another's SM disposition given the virtue of translucency, thereby excluding the SM from cooperative interaction. [29] My argument is that C-S's will, ceterus paribus, be able to obtain cooperative benefits that are not only unavailable to SM's but also unavailable to CM's.

What utility can two C-S's in a PD expect if we assume that translucency cannot detect the latent tendency for defection? First assume that only one C-S can defect without his future opportunities for cooperation being adversely affected. In this

---

[29] Morals By Agreement; op cit., pp. 15, 170, 173.

case the C-S whose future opportunities for cooperation will not be adversely affected will base his actions on a perceived joint strategy and then defect at the opportune moment, exploiting the other C-S. In this case the C-S who defects could expect a greater utility than could either a CM or an SM.

If we now assume that both C-S's could defect without their future opportunities for cooperation being adversely affected then each could only expect the same utility as could an SM. Both would base their actions on a perceived joint strategy and both would defect at the opportune moment. In this case they would be behaviorally indistinguishable from SM's but dispositionally different.

If we now assume that translucency can detect this latent tendency for defection on the part of those persons with C-S dispositions then this raises another issue. To say that translucency can detect this latent tendency for defection is not to say that it can detect whether or not one is actually going to defect on this particular occasion. To say that translucency can detect whether or not a C-S is going to defect on this particular occasion, or on any particular occasion, would be to make translucency the same as transparency. Given that we have already examined the argument for transparency we will assume that translucency can only detect a latent tendency for defection.

Given this assumption what utility could C-S's expect in a PD? If translucency can only detect another's latent tendency for defection rather than whether or not one is actually going to defect on this particular occasion then each person in the dilemma would know that the other person would defect if doing so would not adversely affect his future opportunities for cooperative interaction. Given this assumption there seem to be two possible responses open to the C-S's: rather than take the chance of being exploited they could base their actions on an individual strategy and not cooperate like the SM, or they could take the chance of being exploited (knowing that the opportunities of exploitation are rare) and base their actions on a joint strategy like the CM. The choice they would make in this matter would seem to depend on the ratio of the cost of exploitation over their expected benefit from cooperating. If the cost of being exploited were sufficiently great then it would be rational to base one's actions on an individual strategy and not cooperate. In this case their expected utility would be the same as SM's. If the benefits from cooperation were sufficiently great, such that it offset the costs of occasionally being exploited, then it would be rational to cooperate. In this case their expected utility would be the same as CM's unless one or the other actually were exploited.

We will return to this argument when we come to consider Gauthier's formal argument for choosing a CM disposition and my

formal argument for choosing a C-S disposition, but in the meantime we want to consider the conclusions which can be drawn given each of the assumptions which have been made. In each of the PD arguments Gauthier's position is unchanged regardless of which assumption was made. Given transparency and either version of translucency CM's would cooperate if "given her estimate of whether or not her partner will choose to cooperate, her own expected utility is greater than the utility she would expect from the non-cooperative outcome." [30] SM's never choose to cooperate in PD situations and, given the translucency assumption, can sometimes exploit others if the other mistakes the SM for a CM. C-S's, on the other hand, sometimes cooperate and sometimes don't. In the case where we assumed transparency we saw that sometimes C-S's could have an expected utility equal to that of the CM and, at other times, equal to that of SM's. In the case where translucency could not detect the C-S's latent tendency for defection a C-S could sometimes expect a utility equal to that of the CM, at other times greater than that of a CM and at yet other times equal to that of an SM. Finally, in the case where translucency could detect another's latent tendency for defection it was left undecided what utility she could expect.

We now need to consider the outcomes of these dilemma situations in terms of their Pareto-optimality. In those cases

---

[30] Morals By Agreement; op cit., p. 170.

where we assumed transparency the C-S obtained the Pareto-optimal outcome in every case where the CM obtained the Pareto-optimal outcome. The SM, given his disposition to base his actions on an individual strategy, never obtains the Pareto-optimal outcome. In those cases where we assumed translucency could not detect another's latent tendency for defection the C-S would obtain the Pareto-optimal outcome in every case where the CM obtained the Pareto-optimal outcome except in those cases where one or the other, or both of the C-S's, found that they could defect without their future opportunities for cooperative interaction being adversely affected. Finally, in those cases where translucency could detect another's latent tendency for defection it was left undecided on which occasions the Pareto-optimal outcome would be reached.

Given these arguments we can say on behalf of Gauthier's CM's that CM's can obtain benefits which are unavailable to SM's and CM's can sometimes obtain the Pareto-optimal outcome. We can also say on behalf of my C-S's that C-S's can obtain benefits which are sometimes equal to SM's and CM's and sometimes unavailable to either CM's or SM's and that C-S's can also sometimes obtain the Pareto-optimal outcome. On behalf of the SM's we can say that while they never achieve the Pareto-optimal outcome it is sometimes true that they can obtain benefits by exploiting others which may or may not be unavailable to either of the two alternative dispositions. "Each supposes her

disposition to be rational. But who is right?" [31]

Prior to proceeding with the next chapter, where I attempt to provide an answer to the above-posed question, I have one remaining task to accomplish; I need to argue in favour of my C-S disposition being a more realistic disposition to hold than a CM disposition. Gauthier defines rationality as utility maximization; that is to say, in Gauthier's moral world we are rational insofar as we maximize our utilities. I suggested that Gauthier introduced a false dichotomy when he inferred that CM's and SM's were jointly exhaustive dispositions. I have argued that a C-S disposition is a realistic disposition in Gauthier's moral world on the grounds that, in identifying a need for morality with his argument for the morally free zone, Gauthier left open the possibility of defections occurring thus allowing me to identify and introduce a C-S disposition. However, while a C-S disposition is a realistic possibility in Gauthier's moral world, is it a more realistic conception of what utility maximizers would be? I think it is.

It seems to me that it is more rational to maximize one's expected utilities in all available circumstances than it is to maximize one's expected utilities in all but one available circumstance. This is not an endorsement of an SM disposition but rather a discounting of a CM disposition. A person with a C-S

---

[31] Morals By Agreement; op cit. p. 170.

disposition is maximizing his expected utilities in every available circumstance whereas Gauthier's person with a CM disposition is not maximizing his expected utility in one case where he could. This is especially true given the description of the circumstance where the person with the C-S disposition is maximizing his expected utility and the person with the CM disposition is not.

The person with the C-S disposition is more of an opportunist than the person with the CM disposition, and that is how he should be if he is concerned to maximize his utility. By making condition (iii), when it obtains, a sufficient condition for basing his actions on a perceived joint strategy, a person with a C-S disposition gets to have his cake and eat it too. Like a CM, he is afforded all the opportunities for maximizing his utility that cooperation allows, yet he is afforded the unique and extra opportunity to maximize his utility when doing so will not affect his future opportunities for cooperative interactions. As a rational utility maximizer one would think that a C-S disposition is the more realistic disposition to hold.

We must be cautious to ensure that the distinction between rationality and realism is not denied. Rationality, as Gauthier has defined it, pertains to utility-maximization. Realism, on the other hand, pertains to translucency and its limitations, to the frequency of condition (iii) obtaining and to whether or not

agents would be narrowly or broadly compliant, (an issue which I have not addressed).[32] Realism also pertains to Gauthier's identified need for morality with his morally free zone and the disposition which one would choose given the need for morality.

One question that one might be concerned to raise with respect to Gauthier's CM disposition is: 'why doesn't he take into account condition (iii) as the person with the C-S disposition does when he considers whether or not to base his actions on a joint strategy?' In fact, remember that Gauthier promised us that

> his theory must generate, strictly as principles for choice, and so without introducing prior moral assumptions, constraints on the pursuit of individual interest or advantage that, being impartial, satisfy the traditional understanding of morality.[33]

If he is to accomplish his task without introducing prior moral assumptions then it would seem that Gauthier can only allow for the possibility of a CM disposition after he shows why a C-S disposition is not at least as rational an alternative. By identifying and introducing a C-S disposition I am suggesting that Gauthier's CM is, at times, morally incontinent. My alternative, the person with the C-S disposition, merely takes the occasional "moral holiday" when the opportunity for doing so arises.

---

[32] See Kraus and Coleman, op cit.

[33] Morals By Agreement; op cit., p. 6.

Gauthier's morally incontinent CM thus is "moral" when he need not be and suffers from failing to maximize his utility when he could easily do so, and moreover, do so without suffering any adverse repercussions. If this is true then Gauthier's moral enterprise suffers from a failure to accomplish his designated task without importing some moral premises into his argument. Furthermore, if CM's do suffer from moral incontinence, then my argument in favour of the C-S being at least as rational as the CM gains additional support for C-S's would not really be in direct competition with CM's for they are morally incontinent.

Gauthier might be able to circumvent this objection of moral incontinence on the part of the CM by showing why it is the case that a person with a CM disposition would not defect when this is to his advantage, and including the condition that he would not suffer from any adverse repercussions in terms of being excluded from any future cooperative interactions. If he did manage to show this I think it would be because he redefined what a CM was; a CM would be more like my C-S.

Having said all of this, does this make a person with a C-S disposition a more likely candidate to be found inhabiting the Gauthier moral world? I think so on the grounds that suffering from such a malady as moral incontinence is not a malady that an agent concerned with maximizing his utility would want to be

afflicted with.

In summary I have identified a third alternative to Gauthier's two dispositions, the C-S disposition. I have argued that one can readily distinguish these dispositions, if not on the grounds of their respective behaviors, then on the grounds of their reasoning, i.e., the necessary and sufficient conditions on which they will base their actions. I have also argued that this C-S disposition is at least as rational in some cases as Gauthier's CM disposition. By implication I have suggested that Gauthier has failed to show how he solves the compliance problem given that the person who adopts a CM disposition is morally incontinent. However, the arguments presented in this chapter have really been informal arguments, precursors to the formal argument of Gauthier's when he argues that a disposition of constrained maximization is the uniquely rational strategy. In order to support my position I must redeem some of the conceptual IOU's I have offered throughout this chapter and present a formal argument in favour of a C-S disposition and juxtapose it with Gauthier's formal argument in favour of a CM disposition. This I attempt to do in the following chapter.

Chapter III

The Morally Incontinent CM

Thus far three alternative dispositions, a constrained maximizing disposition (CM), a straightforward maximizing (SM) disposition and a C-S disposition have been identified. Gauthier argues that his CM disposition is the uniquely rational strategy, for a person with a CM disposition would have opportunities to obtain cooperative benefits that are unavailable to SM's and if all rational agents chose a CM disposition the compliance problem would be solved. In terms of the SM disposition, Gauthier argues, that this is not the uniquely rational strategy that Hobbes's Foole thinks it is, for adopting an SM disposition will not generate as much utility overall as would occur if one adopted a CM disposition and it would not solve the compliance problem, thus a political solution would be required.

I argued in the previous chapter that a C-S disposition is at least as rational a strategy in some cases as a CM disposition, on the grounds that a person with a CM disposition suffers from moral incontinence for not taking advantage when he could do so without being discovered, and therefore the person with the C-S disposition would have opportunities to obtain cooperative benefits that are unavailable to those with a CM disposition. I also argued, by implication, that although a C-S disposition does not solve the compliance problem in all cases it

does solve it in any case where a failure to resolve it would result in a loss of future cooperative benefits.

Gauthier, being satisfied with his preliminary arguments and his CM/SM distinction, turns to the general argument of why rational agents would choose a CM disposition over an SM disposition. I will not argue in favour of my C-S disposition in conjunction with Gauthier's argument, thus confusing things, but rather I will first present Gauthier's two arguments, Argument (1) and Argument (2), his final argument, and then I will present my argument. If Gauthier is successful in arguing that a rational agent would choose a CM disposition over an SM disposition, and I think he is -- at least if we allow him his ill-defined notion of translucency --, then I can simply argue that a rational agent would choose a C-S disposition over a CM disposition and ignore the SM disposition.

By ignoring the SM disposition I am not simply running an extension argument on Gauthier's CM disposition. Recall that in the last chapter I argued that one could distinguish a C-S from a CM disposition on the grounds of a behavioral difference and on dispositional grounds as well. CM's took conditions (i) and (ii) as necessary and sufficient conditions for basing their actions on a joint strategy, whereas C-S's took conditions (i) and (ii) as only necessary conditions when defecting would not adversely affect their opportunities for future cooperative interaction and

necessary and sufficient when defecting would adversely affect their opportunities for future cooperative interactions. Moreover, SM's never take conditions (i) and (ii) into account whereas both CM's and C-S's do. C-S's are not simply SM's in their most effective disguise and nor are they CM's. This can be illustrated by an analogy.

Consider the following example of three alternative dispositions which one might hold with respect to promise keeping. First, one might have the disposition to break one's promise whenever it was to one's benefit to do so, regardless if one was discovered or not. Second, one might have the disposition to keep one's promise whenever one believed that the promisee was also likely to keep his promise. Finally, one might have a qualified disposition with respect to promise keeping such that he would keep his promise if he believed that the promisee was also likely to keep his promise and he could not break his promise without being found out.

In some cases, for example, the first disposition and the third disposition when both conjuncts obtained, the behaviour would be the same. Both would make a promise and both would end up breaking it, but there would still be a dispositional difference. In the former case the person would always be expected to break his promise but in the latter case the person would keep his promise if he thought he would be found out if he

broke it and he expected the promisee to also keep his promise. In the case of the second disposition and the third disposition, when only the first conjunct obtained, there would be no behavioral difference but there would be a dispositional difference. The person with the third disposition would be ready to break his promise, if he knew he would not be discovered, whereas the person with the second disposition would still keep his promise even if he knew he could break it without being discovered. If we dissected their respective psychologies we would find that each had a distinct disposition.

Prior to turning to Gauthier's argument it is important to note that a rational agent's choice of disposition is not a case of strategic choice; rather, Gauthier says, it is a case of parametric choice, the same as it was for CM's, SM's and C-S's when they were reasoning parametrically about whether to adopt a joint strategy, a perceived joint strategy, or an individual strategy.

> Taking other's dispositions as fixed, the individual reasons parametrically to his own best disposition. Thus he compares the expected utility of disposing himself to maximize utility given other's expected strategy choices, with the utility of disposing himself to cooperate with others in bringing about nearly fair and optimal outcomes. [1]

Given my C-S disposition a rational agent needs to consider also what his expected utility would be, given not only the utility of

---

[1] Morals By Agreement; op cit., p. 171.

disposing himself to cooperate with others in bringing about nearly fair and optimal outcomes but also what his expected utility would be if defecting would not adversely affect his opportunities for future cooperative interaction. He has to make calculations for his expected utility by considering whether everyone defected, one person defected or both people defected. He need not consider the case where condition (iii) never obtained for I have assumed that there is at least a possibility of condition (iii) obtaining. If this possibility were denied then this would involve denying the possibility of a C-S disposition and Gauthier's identified need for morality, via his morally free zone, would be brought into question.

When faced with the choice of either a CM disposition or an SM disposition, Gauthier argues, the rational agent need only consider those cases where his choice would yield different behaviors. This seems to be perfectly acceptable if there are only two dispositions to choose from, for there is really no choice to be made if the resulting behaviour would be the same and the expected utility would also be the same. However, since I have introduced a third disposition the rational agent must also consider those cases where, even though the behaviour may be the same, the expected utilities would be different. —

When faced with the decision to choose either a CM disposition or an SM disposition, Gauthier argues, the rational

agent only considers those cases where the disposition of constraint would lead her to base her action on a joint strategy and those cases where the disposition to straightforwardly maximize would lead her to base her action on an individual strategy. In order to take into account the possibility of a C-S disposition the rational agent would also have to consider first those rare cases when defecting at the opportune moment would be available for himself only, those cases where defection was available for the other and not himself, and those extremely rare cases where defection was available for everyone. These last cases must be extremely rare for if they were not, as I have already argued, it might not be rational to adopt any kind of cooperative strategy. The rational agent does not need to consider the possibility of condition (iii) never obtaining, for there is a possibility that it will obtain. When considering which disposition to choose Gauthier imposes two necessary conditions:

> First, they must afford the prospect of mutually beneficial and fair cooperation, since otherwise constraint would be pointless. And second, they must afford some prospect for individually beneficial defection, since otherwise no constraint would be needed to realize the mutual benefits.[2]

We now need to consider the detailed arguments Gauthier offers in support of his conclusion that a rational agent would

---

[2] Morals By Agreement; op cit., p. 171.

choose a CM disposition over an SM disposition. (I waive presenting my argument in favour of the C-S disposition until after presenting Gauthier's argument in favour of the CM disposition.) Gauthier presents us with two possible arguments that a rational person would consider.

For each of Gauthier's two arguments there are three possible alternatives the rational agent must consider. First he must consider what his expected utility would be if everyone were to base their action on an individual strategy, i.e., no cooperation by anyone. The second alternative open to the rational agent is what his expected utility would be if all people were to base their action on a joint strategy, i.e., everyone cooperates. The third alternative is that the rational agent bases his action on an individual strategy and the others base their actions on a joint strategy, i.e., he acts straightforwardly and they cooperate.

One thing we need to remind ourselves of, in considering the following arguments, is that the rational agent is not choosing whether to cooperate or not in a particular situation, rather he is deciding on whether to adopt a disposition to cooperate, or not cooperate, given that certain conditions obtain. In effect, he is making a policy decision with respect to his future behaviour. Gauthier considers first an argument in which the

rational agent reasons as follows: [3]

> Argument (1): Suppose I adopt straightforward maximization. Then if I expect the others to base their actions on a joint strategy, I defect to my best individual strategy and expect a utility, u''. If I expect the others to act on individual strategies, then so do I, and expect a utility, u. If the probability that others will base their actions on a joint strategy is p, then my overall expected utility is [pu'' + (1-p)u].
>
> Suppose I adopt constrained maximization. Then if I expect the others to base their actions on a joint strategy, so do I, and expect a utility u'. If I expect the others to act on individual strategies, then so do I, and expect a utility, u. Thus my overall expected utility is [pu' + (1 - p)u].
>
> Since u'' is greater than u', [pu'' + (1 - p)u] is greater than [pu' + (1 - p)u], for any value of p other than 0 (and for p=0, the two are equal). Therefore, to maximize my overall expectation of utility, I should adopt straightforward maximization.

Argument (1), Gauthier says, suffers from an unwarranted assumption. This first argument assumes that the probability of others acting cooperatively is independent of one's own disposition, but the probability is not independent. CM's only act cooperatively if they think the others are also similarly

---

[3] In the following arguments the symbols are identified as follows, where u is less than u' and u' is less than u'':

u - expected utility if each person acts on an individual strategy

u' - expected utility if all act on a joint strategy

u'' - expected utility if the rational agent acts on an individual strategy and others act on a joint strategy

disposed and if they expect the outcome to be nearly fair and optimal; if CM's don't think this then they would base their action on an individual strategy. Therefore, for these two reasons, an SM would not have the opportunities that a CM has; his utility expectation would be less than what he anticipated in Argument (1). Rejecting Argument (1), Gauthier conditionally endorses the following reasoning.

> Argument (2): Suppose I adopt straightforward maximization. Then I must expect the others to employ maximizing individual strategies in interacting with me; so do I, and expect a utility, u.
>
> Suppose I adopt constrained maximization. Then if the others are conditionally disposed to constrained maximization, I may expect them to base their actions on a cooperative joint strategy in interaction with me; so do I, and expect a utility u'. If they are not so disposed, I employ a maximizing strategy and expect u as before. If the probability that others are disposed to constrained maximization is p, then my overall expected utility is [pu' +(1 - p)u].
>
> Since u' is greater than u, [pu' + (1 - p)u] is greater than u for any value of p other than 0 (and for p=0, the two are equal). Therefore to maximize my overall expectation of utility, I should adopt constrained maximization.[4]

Argument (2) does not make the unwarranted assumption found in Argument (1) Gauthier argues. If a CM finds herself amongst other CM's, and expects a nearly fair and optimal outcome, then she will base her action on a joint strategy. On the other hand, if she finds herself amongst a den of SM's, she will not expect a nearly fair and optimal outcome, and therefore will not opt for a

---

[4] Morals By Agreement; op cit., pp. 171-172.

joint strategy. The CM acts differently in those cases where she is amongst a den of SM's and those cases where she is amongst a community of similarly disposed CM's. By doing so, Gauthier argues, she is able to find more opportunities for mutually beneficial cooperation than an SM could. CM's know that SM's would take advantage in cooperative enterprises, and knowing this they would seek to prevent SM's from engaging in joint strategies with themselves, thus limiting the occasions SM's would have for taking advantage. By having more opportunities for cooperative interaction CM's stand to benefit more than SM's.

Gauthier's endorsement of the reasoning in Argument (2) is conditional upon the agents being transparent. That is to say if all appear as they really are, they wear their dispositions on their sleeve, then the CM can expect to do better than the SM. This is unsatisfactory given that the Foole would simply reply that the rational thing to do would be to conceal one's true disposition, thereby gaining the trust and respect of others, and wait for defection opportunities that generated large payoffs.

To circumvent this possible Gauthier has to show how one knows whether he is amongst CM's or SM's. We have seen that Gauthier's answer to this problem lies in his argument for translucency. Translucency, as earlier stated, simply means that one can, "not with certainty, but as more than mere guesswork,"[5]

---

[5] Morals By Agreement; op cit., p. 174.

ascertain the disposition of another in terms of whether he can be expected to cooperate. Gauthier rejects "transparency," i.e., everyone appears in their true colors, in favour of translucency because he thinks attributing transparency to everyone would mitigate any realistic application of his argument. He also rejects beings who are "opaque" for their interactions would result in non-optimal outcomes and thus require a political solution to such problems as parasites and free riders. Moreover, if beings were opaque it may not be rational to agree to cooperative enterprises without some form of external guarantee, substantive rules of conduct enforced by a Leviathan, that compliance would be ensured.

Therefore, assuming that we are translucent to some degree or other, Gauthier offers a third argument to show that constrained maximization is still the more rational strategy.[6] In this third argument non-cooperation, cooperation, defection and exploitation are the four possible outcomes where defection has the value of 1, exploitation has the value of 0, cooperation has a value of less than 1 (but greater than 0), and non-

---

[6] Gauthier makes the following simplifying assumptions: "the non-cooperative outcome results unless (i) those interacting are CM's who achieve mutual recognition, in which case the cooperative outcome results, or (ii) those interacting include CM's who fail to recognize SM's but are themselves recognized, in which case the outcome affords the SM's the benefits of individual defection and the CM's the costs of having advantage taken of mistakenly basing their actions on a cooperative strategy." Furthermore, he ignores the inadvertant taking of advantage when CM's mistake their fellows for SM's. Cf Morals By Agreement; op cit., p. 175.

cooperation also has the value of less than 1 but greater than 0 and furthermore the value of non-cooperation is less than the value of cooperation. [7] Gauthier's new argument for why it is rational to choose a CM disposition over an SM disposition is different from both Arguments (1) and (2). Consider the following:

> Let us now calculate the expected utilities for CM's and SM's in situations affording both the prospect of mutually beneficial cooperation and individually beneficial defection. A CM expects the utility nc unless (i) she succeeds in cooperating with other CM's or (ii) she is exploited by an SM. The probability of (i) is the combined probability that she interacts with a CM, r, and that they achieve mutual recognition, p, or rp. In this case she gains (c - nc) over her non-cooperative expectation nc. Thus the effect of (i) is

---

[7] This is a list of the variables and their assigned values and/or constraints. Gauthier uses the variables u'' for cooperation and u' for non-cooperation. To make things simpler I use 'c' for cooperation and 'nc' for non-cooperation, 'd' for defection and 'e' for exploitation.

$d = 1$
$c < d \ \& \ c > e$
$nc < c \ \& \ nc > e$
$e = 0$

The remaining variables are p, q, and r: the values of p, q, and r fall between 0 and 1

p - the probability that CM's achieve mutual recognition, hence they cooperate

q - the probability that CM's fail to recognize SM's but will themselves be recognized, so that exploitation and defection will result

r - the probability that a randomly selected member of the population is a CM, assuming that everyone is either a CM or an SM, so the probability that a randomly selected person is an SM is (1 - r)

to increase her utility expectation by a value of [rp(c-nc)]. The probability of (ii) is the combined probability that she interacts with an SM, 1-r, and that she fails to recognize him but is recognized, q, or (1-r)q. In this case she receives 0, so she loses her non-cooperative expectation nc. Thus the effect of (ii) is to reduce her utility expectation by a value [(1-r)qnc]. Taking both (i) and (ii) into account, a CM expects the utility (nc + [rp(c - nc)] - (1-r)qnc).

An SM expects the utility nc unless he exploits a CM. The probability of this is the combined probability that he interacts with a CM, r, and that he recognizes her but is not recognized by her, q, or rq. In this case he gains (1-nc) over his non-cooperative expectation nc. Thus the effect is to increase his utility expectation by a value [rq(1-nc)]. An SM thus expects the utility (nc + [rq(1-nc)]).

It is rational to dispose oneself to constrained maximization if and only if the utility expected by a CM is greater than the utility expected by an SM, which obtains if and only if p/q is greater than ((1-nc)/(c-nc) + [(1-r)nc]/[r(c-nc)]).

The first term of this expression, ((1-nc)/(c-nc), relates the gain from defection to the gain through cooperation. The value of defection is of course greater than that of cooperation, so this term is greater than 1. The second term, [(1-r)nc]/[r(c-nc)]), depends for its value on r. If r=0 (i.e., there are no CM's in the population), then its value is infinite. As r increases, the value of the expression decreases, until if r=1 (i.e., there are only CM's in the population) its value is 0.[8]

Gauthier wants to draw two conclusions from this argument. "First, it is rational to dispose oneself to constrained maximization only if the ration of p to q, i.e., the ratio between the probability that an interaction involving CM's will result in cooperation and the probability that an interaction involving CM's and SM's will involve exploitation and defection,

---

[8] Morals By Agreement; op cit. pp. 175-176.

is greater than the ratio between the gain from defection and the gain through cooperation." [9] The second conclusion Gauthier wants to draw is that "as the proportion of CM's in the population increases (so that the value of r increases), the value of the ratio p to q that is required for it to be rational to dispose oneself to constrained maximization decreases. The more constrained maximizers there are, the greater the risks a constrained maximizer may rationally accept of failed cooperation and exploitation. However, these risks, and particularly the latter, must remain relatively small." [10]

Having presented Gauthier's three arguments we now need to ascertain first, what is it that he actually proves with Arguments (1) and (2), and second, do Arguments (1) and (2) prove what Gauthier thinks they prove, and finally, do Arguments (1) and (2) tell us anything which we did not know already? We also need to see whether Gauthier's third argument supports his conclusion that, given translucency, constrained maximization is the more rational disposition to adopt. The question we are most concerned to answer though is whether constrained maximization is

---

[9] Morals By Agreement; op cit., p. 176.

[10] Morals By Agreement; op cit., p. 176. Note Gauthier's last sentence in this quote: i.e., that the risks of failed cooperation and exploitation must remain relatively small. This supports my argument in two ways. First it concedes the argument I made about Gauthier's morally free zone allowing for possible cases of defection and exploitation thus allowing me to introduce the C-S disposition. Second, this supports my argument that the cases of exploitation and defection must be rare.

a more rational disposition to adopt than a C-S disposition, or, are there at least some cases where it would be more rational to adopt a C-S disposition rather than a CM disposition?

Gauthier says, with respect to Arguments (1) and (2), "what we have shown is that, if the straightforward maximizer and the constrained maximizer appear in their true colors, then the constrained maximizer must do better." [11] If this is all Arguments (1) and (2) show then it is of little importance. We already know that, for example in a two-person one-shot Prisoner's Dilemma, people who cooperate can expect to do better than people who do not cooperate! The problem, which Gauthier has failed to answer with these two arguments, is how it can be the case that people with competing interests can achieve the optimal outcome given the conditions A, B, and C of strategically rational choice. By simply saying, in essence, "constrain yourself and act so that the outcome which you mutually achieve is the optimal outcome and you will be rewarded for you efforts afterwards" Gauthier has conflated their competing interests into the mutual interest of how both can get the best out of the situation in which they presently find themselves. In conflating their interests Gauthier may have denied his premise of non-tuism, i.e., their not taking an interest in the other's interest. Gauthier's Arguments (1) and (2) do prove what Gauthier says they prove but his conclusion is trivial.

---

[11] Morals By Agreement; op cit., p. 173.

Arguments (1) and (2) are cases of parametric reasoning, Gauthier says. That is to say they are cases of an agent who takes others' dispositions to be fixed and then reasons to his own best disposition. Gauthier's third argument, on the other hand, is not a case of parametric reasoning. On the contrary it seems to be a case of finding "the conditions under which the decision to dispose oneself to constrained maximization is rational for translucent persons," and asking "if these are (or may be) the conditions in which we find ourselves." [12] This is an entirely different kind of argument than either Argument (1) or Argument (2).

In the first two arguments Gauthier was concerned to show that the disposition of constrained maximization yielded a greater utility expectation than did a disposition of straightforward maximization. In his third argument he is seeking the "conditions under which the decision to dispose oneself to constrained maximization" would be rational. By adopting this alternative approach Gauthier leaves the door wide open for my argument in favour of a C-S disposition. I too can seek the conditions under which the decision to dispose oneself to a C-S disposition would be rational. I could also ask if these are (or may be) the conditions in which we find ourselves but I ignore this question given the ideal nature of the agents under

---

[12] Morals By Agreement; op cit., p. 174.

consideration.

By saying that he is seeking for the conditions under which the decision to dispose oneself to constrained maximization would be rational Gauthier is implying first that there may be cases where it would not be rational to dispose oneself to a CM disposition. This is surprising for even if one were in a society which consisted of only SM's the CM disposition leaves open the possibility of acting straightforwardly when one expects to be exploited by another. One does not give up a CM disposition by acting straightforwardly in these circumstances for acting straightforwardly is an inherent part of the CM disposition. The second thing implied by Gauthier's claim is that a CM disposition may not be the uniquely rational strategy; this would depend on which conditions the rational agent found himself in. I am not concerned to argue that a C-S disposition is the uniquely rational strategy, rather, my concern is to argue for the weaker claim that there are conditions under which it would be rational to choose a C-S disposition.

There are problems with Gauthier's third argument which he does not address. What do CM's do, given the assumption of translucency, when they do not know what disposition the other person holds? Recall that in the previous chapter where I was concerned to see what C-S's would do in a PD I concluded that there were two alternatives open to the agent if he did not know

what disposition the other held.[13] In these kind of cases there are four possible outcomes in terms of recognizing another, call them r1 to r4. In r1 we have mutual recognition; in r2, I recognize him but he does not recognize me; in r3, he recognizes me but I do not recognize him; and in r4, we are unable to recognize each other. Gauthier needs to provide an account of what his CM would do in each of these four situations for, by assuming translucency he leaves himself open to any of these possibilities occurring. His third argument only takes into account r1 and r3. By neglecting r2 and especially r4 Gauthier's argument in favour of CM's is invalid.

One could say that Gauthier implicitly takes r2 into account for if the CM recognized the SM and the SM did not recognize her then the CM would not cooperate for she would not expect an outcome which was nearly fair and optimal. This may be the case but it will not mitigate the damage done to his argument by failing to take into account r4. R4 is crucial for we need to

---

[13] First, rather than take the chance of being exploited they could base their actions on an individual strategy and not cooperate like the SM, or second, they could take the chance of being exploited (knowing that the opportunities of exploitation are rare) and base their actions on a joint strategy like the CM. The choice they would make in this matter would seem to depend on the ratio of the cost of exploitation over their expected benefit from cooperating. If the cost of being exploited were sufficiently great then it would be rational to base one's actions on an individual strategy and not cooperate. In this case their expected utility would be the same as SM's. If the benefits from cooperation were sufficiently great, such that it offset the costs of occasionally being exploited, then it would be rational to cooperate. In this case their expected utility would be the same as CM's unless one or the other actually were exploited

know, not what the SM will do when he doesn't recognize the other -- we already know he won't cooperate regardless of what disposition the other has -- but rather what will the CM do in those cases when she doesn't recognize the other.

When I considered my C-S disposition in the same situation I conceded that there were two possibilities. Suppose Gauthier were to adopt the first alternative -- rather than take the chance of being exploited a CM could base her actions on an individual strategy and not cooperate like the SM -- then what? If he adopted this alternative then not only would he have to revise his third argument but the CM's expected utility in this third argument would need to be revised; it might not be as great as Gauthier thinks it would be. Furthermore, if Gauthier opted for this first alternative then he would have to revise both of the two conclusions he wanted to draw so that he could accommodate those cases where another's disposition was unknown.

Suppose he were to adopt the second alternative -- a CM could take the chance of being exploited assuming that the opportunities of exploitation are rare, (but this depends on the distribution of CM's and SM's in the population), and still base her actions on a joint strategy -- then what? If this were the case then, again, the same three things follow; his argument would need to be revised, therefore the expected utility would be different, and the two conclusions drawn from this argument would

have to be readdressed.

Another problem with Gauthier's third argument is the epistemological problem identified earlier: they may be mistaken in their perceptions and/or their apprehensions of a given situation and therefore it is not the case that a CM expects the utility (nc + [rp(c - nc)] - (1-r)qnc) but rather her expected utility is the probability of (nc + [rp(c - nc)] - (1-r)qnc).

One could expect someone to object to this so-called epistemological problem on the grounds that Gauthier's argument is an ideal argument and he doesn't address epistemological concerns. However, this third argument is Gauthier's formal argument in favour of a CM disposition and he wants to tie this argument to the real world (remember this argument presupposes translucency and translucency was the tie to the real world) and therefore this counter claim will not suffice to overcome this objection. Given that these criticisms are sufficient to render Gauthier's third argument invalid, especially the criticism that he fails to account for what CM's would do if they did not recognize the other, I can now turn to my argument.

Given the ideal nature of Gauthier's arguments for a CM disposition it is justified if I make some simplifying and idealized assumptions in order to present my argument in favour

of the C-S disposition.[14] Therefore, in the first part of this argument, I assume two simple societies, the first society consisting of only CM's and the second society consisting of only C-S's. [15] I also assume that the agents involved know the length of their lives such that they know the number of cooperative opportunities they can expect. The next assumption in this first argument is that the agents in the CM society always cooperate. The final assumption is that the agents in the C-S society always cooperate unless they are in a defect mode.

The question, given these assumptions, is what is the expected utility for a CM and a C-S? In order to calculate their expected utilities I make the following assignments: k - number of cooperative opportunities, x - number of occasions a C-S is in a defect mode, d - defection, c - cooperation, nc - non-cooperation, e - exploitation and d > c, c > nc, nc > e. It is further stipulated that the values of c and nc fall between 1 and 0 and that d=1 and e=0.

Having said this a CM in a society consisting of only CM's has an expected utility of k(c): the number of opportunities for cooperation over her cooperative expectation. A C-S on the other

---

[14] I am indebted to Steven DeHaven for his assistance in formulating the two arguments which follow.

[15] From now on I will ignore the SM disposition assuming that Gauthier's argument in favour of the CM disposition won over the argument in favour of the SM disposition.

hand, in a society of only C-S's, would have an expected utility of [(k-x)*((k-x)/k)*c] + [(k-x)*(x/k)*e] + [x*((k-x)/x)*d)] + [x*(x/k)*nc]. The first term of this expression, [(k-x)*((k-x)/k)*c], calculates the utility a C-S can expect if he is interacting with another C-S who is not in the defect mode. The second term of this expression, [(k-x)*(x/k)*e], calculates the utility a C-S can expect if he is interacting with another C-S who is in the defect mode. The third term of this expression, [x*((k-x)/x)*d)], calculates the utility a C-S could expect if he were in the defect mode and the other was not. The fourth term of this expression, [x*(x/k)*nc], calculates the utility a C-S could expect if they achieved the non-cooperative outcome, i.e., they were both in the defect mode.[16]

In this case of simple societies where the first society is inhabited only by CM's who always cooperate and the second society is inhabited only by C-S's who always cooperate unless they are in a defect mode we find that there are conditions under which the decision to dispose oneself to a C-S disposition is rational. That is to say there are conditions under which choosing a C-S disposition would yeild a greater expected utility than choosing a CM disposition. This is not to say that we cannot, or Gauthier cannot, find cases, given these same assumptions but assigning different values to the variables,

---

[16] See Appendix B, the first problem, for the calculated solutions to both CM and C-S such that the expected utility of a C-S is greater than the expected utility of a CM.

where it would be rational to choose a CM disposition rather than a C-S disposition. [17]

Consider now a mixed society of only CM's and C-S's. In this case we need to know the ratio of CM's to C-S's in the population, therefore, we assign the variable r1 to the number of CM's in the population and (1-r1) for the number of C-S's. The remaining variables are unchanged, i.e, as before, CM's always cooperate and C-S's always cooperate unless they are in the defect mode.

A CM has the following utility expectation in a mixed society of only CM's and C-S's. CM = [k*(1-r1)*(x/k)*e] + [k*r1*c] + [k*(1-r1)*((k-x)/k)*c]. The first term of this expression, [k*(1-r1)*(x/k)*e], calculates the utility a CM can expect if he is exploited by a C-S. The second term of this expression, [k*r1*c], calculates the utility a CM can expect if he cooperates with another CM. The third term of this expression, [k*(1-r1)*((k-x)/k)*c], calculates the utility a CM can expect if he interacts with a C-S who is not in a defect mode.

A C-S has the following utility expectation in a mixed society of only CM's and C-S's. C-S = [(k-x)*(x/k)*(1-r1)*e] + [(k-x)*r1*c] + [(k-x)*(1-r1)*((k-x)/k)*c] + [x*(r1*d)] + [x*(1-

---

[17] See Appendix B, the second problem, where I found values to show that it would be rational to choose a CM disposition rather than a C-S disposition.

r1)*((k-x)/k)*d] + [x*(1-r1)*(x/k)*nc]. The first term of this expression, [(k-x)*(x/k)*(1-r1)*e], calculates the utility a C-S can expect if he is exploited by another C-S who is in the defect mode. The second term of this expression, [(k-x)*r1*c], calculates the utility a C-S could expect if he were not in a defect mode when interacting with a CM and thus they cooperated. The third term of this expression, [(k-x)*(1-r1)*((k-x)/k)*c], calculates the utility a C-S could expect if he interacted with another C-S who, like himself, was not in the defect mode. The fourth term of this expression, [x*(r1*d)], calculates the utility a C-S could expect if he interacted with a CM while he was in a defect mode. The fifth term of this expression, [x*(1-r1)*((k-x)/k)*d], calculates the utility a C-S could expect if he were in a defect mode and he interacted with another C-S who was not in a defect mode, thereby gaining the benefits of defection over this other C-S. The sixth term of this expression, [x*(1-r1)*(x/k)*nc], calculates the utility a C-S could expect if he was in the defect mode and he interacted with another C-S who was also in the defect mode.

Once again we can conclude that even in a mixed society of only CM's and C-S's there are cases where it is rational to choose a C-S disposition.[18] Having said this though it remains

---

[18] See Appendix B, the third problem, where I found values to show that it was rational to choose a C-S disposition over a CM disposition in this mixed society of CM's and C-S's. I did not want to make the strong claim that it was always rational to choose a C-S disposition, given the assumptions that have been

for me to point out that my argument in favour of the C-S disposition is somewhat different than Gauthier's any of Gauthier's three arguments. Gauthier's first two arguments were cases of parametric reasoning such that an agent reasoned to his own best disposition given transparency. Gauthier's third argument was a case of finding the conditions under which it would be rational to dispose oneself to constrained maximization if people were translucent. My arguments are similar to Gauthier's first two arguments in that they are cases of parametric reasoning and my arguments are similar to Gauthier's third argument in that I too looked for conditions under which it was rational to choose a C-S disposition. In neither of my two arguments did I rely on translucency or on transparency but rather I calculated the utility expectation given that a person would interact with another who held a given disposition.

My first argument can best be described as a case of parametric reasoning in which, taking the other's disposition as fixed, i.e., all CM's in the society or all C-S's in another society, the rational agent asks what would be the rational disposition to choose if one were to have a choice of living in

---

made, for I thought there would be cases where it was rational, under these circumstances, to choose a CM disposition. However, in performing these calculations I was unable to generate any values, keeping the imposed constraints in mind, that would make it rational for one to choose a CM rather than a C-S disposition. Recall that in the case of the two simple societies I was able to generate such a case.

either the first society, the all CM society, or living in the second society, the all C-S society? Given that there were cases where it was more rational to be a C-S and given that there were also cases where it was more rational to be a CM, (this depended on what values were assigned to cooperation and non-cooperation), then Gauthier's claim that constrained maximization is the uniquely rational strategy is called into question while my claim that a C-S disposition is at least as rational a strategy in some cases is vindicated.

The second argument I presented was also a case of parametric reasoning about what disposition it would be rational to adopt given that the agent was going to enter into a mixed society of only CM's and C-S's. Contrasting this argument with the first argument I presented I was unable to generate any value for cooperation and non-cooperation, (keeping of course within the imposed constraints), such that it made it rational for one to choose a CM disposition. This second argument was similar to the first in that neither presupposed translucency nor transparency.

Given that there are conditions under which it would be rational to chose a C-S disposition as opposed to a CM disposition I can conclude that Gauthier's constrained maximizer is, at times, morally incontinent. Furthermore I can conclude that constrained maximizing is not the uniquely rational strategy

Gauthier thinks it is. This should not surprise us for surely it is in accord with some of our intuitions, i.e., if one can take a moral holiday without adversely affecting his future opportunities for cooperation then that person can expect a greater utility than the person who doesn't take the moral holiday when he could do so.

There are still two areas of concern which I need to address. The first is, will the Ideal Actor from the Archimedean Point endorse the C-S disposition? The second area of concern is the symmetry problem. That is to say I have to argue that all rational agents who find themselves in certain conditions will choose a C-S disposition.

In regards to the Ideal Actor I think he would endorse a C-S disposition. Recall that the Ideal Actor is to choose "not as if she had an equal chance of being each person affected by her choice, but as if she were each of these persons." [19] Being concerned to maximize her utilities the Ideal Actor would consider the possibility of a C-S disposition. Insofar as she found that a C-S disposition lead to a greater utility expectation than did a CM disposition she would choose the former. Of course it need not be the case that she would always choose the C-S disposition. Presumably one could construct an argument, given different constraints, that would show it was

---

[19] Morals By Agreement; op cit., p. 255.

rational to adopt some other disposition, be it a CM disposition or even some entirely different disposition. This however need not concern us for I have made my case.

The second point concerning the Ideal Actor's choice is that he would have to know not only that the opportunities for defecting from a perceived joint strategy are sufficiently rare such that they do not make it irrational to choose a cooperative disposition, be it a CM disposition or a C-S disposition but also that the cases where both agents were in a defect mode were rarer than the cases of one agent being in a defect mode.

The answer to the symmetry problem lies in Gauthier's identified need for morality. If the world were a perfectly competitive market, Gauthier argues, there would be no need for morality. The world isn't a perfectly competitive market, given the problems posed by free riders and parasites, and the world doesn't usually match our idealized conceptions of it. Opportunities to defect from perceived joint strategies are a realistical possibility in an imperfect world, and rational agents would do well to take heed of such possibilities.

They can take heed of such possibilities if they account for them when they make their calculations concerning which disposition they will have. I granted that if such possibilities were frequent in number then adopting a C-S disposition might not

be rational for anyone; but then adopting a CM disposition might also not be rational for anyone. Rational agents would be aware that the possibility of being the exploited and being the exploiter existed, rare though the case may be. If such possibilities were sufficiently rare then it may be the case that the chance of being exploited would be offset by the benefits gained from cooperative interaction. The benefits of cooperative interaction are great, as Gauthier testifies with his appraisal of modern western society and its accomplishments. In order for rational agents to conclude that a C-S disposition was the rational strategy to adopt, given the three identified alternatives, they would have to assess whether the benefits of cooperative interaction with the chance of occasional exploitation were greater than the benefits offered by the political solution of an inevitable Leviathan. I think they would find the benefits of cooperative interaction and the freedom associated with such voluntary interaction much preferable to the war of "all against all," where life would be "nasty, brutish and short."

Having said all of this I can only conclude that Gauthier failed in his attempt to derive morality from an instrumental conception of rationality. The reasons for his failure have been argued for throughout this paper. In some ways my arguments, though critical, may help to strengthen Gauthier's moral enterprise. I suggest that some version of a C-S disposition,

although not necessarily the version I have offered, may be identified and substituted in place of a constrained maximizing disposition.

Bibliography

Baier, Kurt; "Rationality, Value and Preference," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Braybrooke, David; "Social Contract Theories' Fanciest Flight," Ethics; Vol. 97, No. 4, July 1987.

Buchanan, James; "The Gauthier Enterprise," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Fishkin, James; "Bargaining, Justice and Justification: Towards Reconstruction," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Gauthier, David; "Reply to Wolfram," Philosophical Books, Vol. 28, No. 3, July 1987.

Gauthier, David; "A Response," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Gauthier, David; Morals By Agreement; Clarendon Press, Oxford University Press, Great Britain; 1986.

Gibbard, Allan; "Reasonably Reciprocal, David Gauthier's Morals By Agreement;" Times Literary Supplement, February 20, 1987.

Hare, R.M.; Moral Thinking: Its Levels, Method and Point; Clarendon Press, Oxford University Press; Great Britain; 1981.

Hardin, Russell; "Bargaining for Justice," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Harman, Gilbert; "Rationality in Agreement," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Kraus, Jody S. and Coleman, Jules L; "Morality and the Theory of Rational Choice," Ethics; Vol. 97, No. 4, July 1987.

Luce, Duncan R. and Raiffa, Howard; Games and Decisions: Introduction and Critical Survey; John Wiley and Sons, Inc., New York; 1958.

Mendola, Joseph; "Gauthier's Morals By Agreement and Two Kinds of Rationality," Ethics; Vol. 97, No. 4, July 1987.

Morris, Christopher; "The Relation Between Self-Interest and Justice in Contractarian Ethics," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Narveson, Jan; "David Gauthier, Morals By Agreement;" Canadian Philosophical Reviews, Vol. 8, No. 7, July 1987, pp. 269-272.

Parfit, Derek; Reasons and Persons; Oxford University Press; Great Britain; 1986.

Rawls, John; A Theory of Justice; The Belnap Press of Harvard University Press, Cambridge, MA; 1971.

Sumner, L.W.; "Justice Contracted;" Unpublished review of Morals By Agreement; University of Toronto; March 31, 1987.

Sumner, L.W.; "Justice Contracted;" Dialogue; Vol XXVI, No. 3, Autumn, 1987.

Suppes, Patrick; Axiomatic Set Theory; Dover Publications, New York, 1972.

Thomas, Laurence; "Rationality and Affectivity," Unpublished paper presented at the Morals By Agreement Conference in Bowling Green, Ohio in April 1987.

Vallentyne, Peter; "Gauthier on Rationality and Morality," Eidos; Vol. V, No. 1; June 1986.

von Mises, Ludwig, Human Action; Contemporary Books, Inc., Chicago; 1963.

Webster, Michael; "Minimax Concession," Eidos; Vol. V, No. 1; June 1986.

Wolfram, Sybil; "Morals By Agreement," Philosophical Books, Vol. 28, No. 3, July 1987.

Appendix A

Brian and Jane have a problem, the same problem we all face once we are each concerned to act in our own best interest and where the outcome which best manifests that interest is directly dependant on what someone else does. Gauthier must resolve this problem of strategic interaction and he must do so within the confines of the idealizing conditions A, B and C.

Having outlined the nature of the problem we now turn to the first step of his argument.[1] The first step is to enlarge the scope of choice to include lotteries with actions as prizes.[2] There are several important terms in this claim which need to be defined. An "action" is "the object of a choice." [3] There are two types of lotteries; the first type of lottery is one where an action is performed under risk or uncertainty with possible outcomes as prizes. The second type of lottery is a lottery concerning choice with possible actions as prizes. In answering the problem of strategically rational choice Gauthier is going to

---

[1] I do not attempt to evaluate Gauthier's formal apparatus in this Appendix instead I leave this task for those who are interested in game theoretic approaches to bargaining.

[2] Morals by Agreement, op cit., p. 65. The reason Gauthier wants to extend their choices to include lotteries with actions as prizes is so that he can show that Jane and Brian can arrive at an equilibrium outcome via a mixed strategy. Note, an expected outcome is in equilibrium iff it is the product of mutually utility maximizing strategies. See p. 65 in Morals by Agreement.

[3] Morals by Agreement, op cit., p. 62.

use the second type of lottery, the one with possible actions as prizes.

A "strategy" is "a lottery over possible actions." [4] There are two types of strategies; the first type of strategy is a pure strategy. This is where one assigns the probability of 1 to one action and 0 to all other actions. This type of strategy has only one prize and the prize is always awarded. The second type of strategy is a mixed strategy; it is a lottery with several actions as prizes. In a mixed strategy one assigns a non-zero probability to more than one action with the sum of probabilities assigned being equal to 1.

Gauthier makes two innocent idealizing assumptions with respect to lotteries. The two assumptions are that there is "universal availability of a randomizing device" and that this device is "costless to use." [5] It should be noted as well that Gauthier, by expanding the scope of choice from actions to strategies, has not increased the range of possible outcomes. This is the case because strategies are simply lotteries over possible actions and regardless of the outcome of the lottery

---

[4] Morals by Agreement, op cit., p. 65.

[5] Morals by Agreement, op cit., p. 65. Although this assumption about the random device being costless to use is an innocent assumption as an idealizing assumption in strategic rational choice it may not be so innocent in partial compliance theory. This issue is raised by Kraus and Coleman in Ethics, Volume 97, Number 4, July, 1987, in an article entitled "Morality and Rational Choice."

there is still only one possible action. This is especially true for pure strategies as there is but one prize for a pure strategy.

Having introduced the concept "outcome," [6] Gauthier goes on to define an "expected outcome." [7] An expected outcome is simply the "product of the lotteries or strategies chosen by each person." He then introduces the notion of "Cartesian product"[8] saying that "the set of expected outcomes is thus the product of the actor's sets of strategies." This is to say that the product of the lotteries or strategies chosen by each person is the Cartesian product of the actors' sets of strategies.

Outcomes can be in equilibrium or not. An outcome is in

---

[6] An "outcome, as defined with respect to parametric choice, is "the product of an action and a set of determinate circumstances," but in strategic choice an outcome results from "the choices of several actors," ... "one for each person involved in interaction." (See Morals By Agreement; p. 62.)

[7] Morals by Agreement, op cit., p. 65.

[8] A "Cartesian product" is defined as follows: the cartesian product of two sets A and B (in symbols: A * B) is the set of all ordered pairs (x,y) such that x is a member of A and y is a member of B. For example, if

A = {1,2}
B = {Archimedes, Eudoxus}

then

A * B = {(1, Archimedes), (1, Eudoxus), (2, Archimedes), (2, Eudoxus)}

See Suppes, Patrick; Axiomatic Set Theory; Dover Publications, Inc., New York, 1972; p. 49.

equilibrium if and only if (iff) it is the "product of strategies, each of which maximizes the expected utility of the person choosing it given the strategies chosen by the other persons," or to put it in other words, "an outcome is in equilibrium iff it is the product of mutually utility maximizing strategies." [9] Furthermore, an outcome can be in either "strong" or "weak" equilibrium. An outcome is in strong equilibrium iff "each strategy is the actor's only utility maximizing response to other strategies." [10] An outcome is in weak equilibrium if this condition is not met; that is to say an outcome is in weak equilibrium if there is more than one strategy which is a utility maximizing response to other strategies.

Applying these concepts to the Jane and Brian case we find that given a pure strategy none of the outcomes are in equilibrium. They are not in equilibrium because none of the strategies open to Jane and Brian maximizes their utility given the strategy chosen by the other. In order for an outcome for Jane and Brian to be in equilibrium they must choose a mixed strategy. They must choose a mixed strategy as this will afford them a greater number of possible outcomes, one of which may be in strong equilibrium.

However one cannot have a mixed strategy unless one has an

---

[9] Morals by Agreement, op cit., p. 65.

[10] Morals by Agreement, op cit., p. 70.

"interval measure of preference." [11] Given that when we make a choice under risk or uncertainty we must choose between actions, and not outcomes, and given that the actions associated with these outcomes are only probable, then outcomes are also only probable. An interval measure is, in this context, the probability of an action (such action having an associated outcome which is also probable). An aid for determining interval measures are indifference questions. If we can ascertain the indifference ratios of the participants in the interaction then we can assign utilities to these preferences. [12]

---

[11] An "interval measure" is to be distinguished from an "ordinal measure." An ordinal measure is just a hierarchical ranking of preferences for outcomes given the utility of the person for any given outcome. Such a ranking of preferences occurs very simply and quite frequently in choice under certainty. However, in choice under uncertainty it is not simply enough to rank preferences for outcomes, and the associated utilities, because the outcomes are not certain and therefore the utilities associated with those outcomes are also not certain. To circumvent this problem one must make choices not just on the utilities of possible outcomes but also on the probabilities of those outcomes. See p. 24 and p. 42 in Morals By Agreement. Also see my explanation of "Cartesian product" where the interval measure of preference are those possibilities for C and NC which fall between 0 and 1.There are some formal conditions for defining "interval measures" for rational choice theories. (See Luce and Raiffa; pp. 23-31.)

[12] It should be noted that preferences must be both coherent and considered. One's preferences are considered preferences "if and only if there is no conflict between their behavioral and attitudinal dimensions and they are stable under experience and reflection." See Morals By Agreement p. 32. Gauthier wants to argue that if there is a conflict between one's revealed and expressed preferences then the agent lacks an "adequate basis for rational choice" and therefore the agent would not be acting rationally if she made a choice on the basis of these conflicting preferences. See Morals By Agreement p. 28. To say that one's preferences are "coherent" is just to say that they are logically congruent.

Using Gauthier's example [13] where he assigns utilities of 1, 2/3, 1/3 and 0 to the four outcomes in order of Jane's preferences and 1, 1/2, 1/6 and 0 to the four outcomes of Brian's preferences we have the following matrix.

		Brian: Go	Brian: Stay
Jane	Go	0, 1	1, 0
	Stay	2/3 1/6	1/3, 1/2

While we will not reiterate Gauthier's calculations showing Jane's expected utilities if she assigns a probability of 1/4 to going to the party and 3/4 to staying at home we will verify Gauthier's claim that assuming that Brian chooses a mixed strategy wherein he assigns a probability of 1/2 to going to the party and 1/2 to staying at home that Jane's expected utility, whatever she chooses, is 1/2.

Thus if Jane chooses J4 we have [(0*1/2)+(1*1/2)] equals 0 + 1/2 which equals 1/2. If J3 then [(2/3*1/2)+(1/3*1/2)] equals 2/6 + 1/6 which equals 3/6 or 1/2. If J2 then [2/3*1/2)+(1/3*1/2)] equals 2/6 + 1/6 which equals 3/6 or 1/2. If J1 then [(0*1/2)+(1*1/2)] equals 0 + 1/2 which equals 1/2.

---

[13] Morals by Agreement, op cit., p. 66-67.

It would seem that Gauthier is correct; regardless of which preference Jane chooses, if Brian chooses a mixed strategy of assigning 1/2 to going to the party and 1/2 to staying at home, then Jane's expected utility will be 1/2. [14] Furthermore, each outcome is in equilibrium because it is a utility maximizing response given the other's choice but, because it is not the only utility maximizing response, it is not in strong equilibrium. Note that in this example conditions A, B, and C have also been satisfied. Condition A has been satisfied because Jane's response is a rational response (i.e., utility maximizing response) to the choices she expects Brian to make; condition B has been satisfied because Jane expected Brian to satisfy condition A (which he did) and Brian expected Jane to satisfy condition A (which she did); condition C is satisfied as well because both Jane and Brian expected their choices and expectations to be reflected in the expectations of each other (which they were).

Gauthier draws two conclusions concerning strategically rational choice from this example, both of which rest on defining rational response as a utility maximizing response. The first conclusion he draws is that "there must be at least one set of mutually utility maximizing strategies." [15] The second conclusion drawn is that "each person should relate his choice of strategy

---

[14] It should be noted that one really only had to do two calculations of Jane's preferences because Brian's mixed strategy of 1/2 and 1/2 effectively give Jane only two options.

[15] Morals by Agreement, op cit., p. 68.

and expectations of other's choices to a set of mutually utility maximizing strategies." [16]

Having defined a rational response as a utility maximizing response Gauthier is left with the problem of having outcomes which are only in weak equilibrium, i.e., there is more than one utility maximizing response to the other's strategy, rather than strong equilibrium where there is only one utility maximizing response to the other's strategy. To circumvent this potential problem Gauthier introduces the concept of a "centroid utility maximizing response." A centroid utility maximizing response is the rational response one makes to the choices one expects the other to make where this "the rational response" is defined as the response "determined by lottery giving equal probability to all responses satisfying other rationality requirements." [17] The centroid utility maximizing response must be determined by lottery because if it were not then one would, in effect, be revealing one's preference for a given outcome. This would entail that person assigning a greater expected utility to that expected outcome than to the other expected outcomes.

If one were to assign a greater expected utility to one expected outcome over another expected then two assumptions would be violated. The first assumption that would be violated is the

---

[16] Morals by Agreement, op cit., p. 86.

[17] Morals by Agreement, op cit., p. 69.

assumption that each expected outcome be equally utility maximizing. The second assumption that would be violated is the assumption that preference be revealed in rational choice.

Gauthier uses the concept of centroid utility maximizing response to avoid having these two assumptions violated. [18] This in turn requires a reformulation of one of the two conclusions drawn from defining a "rational response" in condition A as a utility maximizing response. We saw earlier that Gauthier wanted to draw the conclusion, concerning strategically rational choice, that "each person should relate his choice of strategy and expectations of other's choices to a set of mutually utility maximizing strategies." [19] Now, in order to avoid violating the assumption of preferences being revealed in rational choice and the assumption that strategies be equally utility maximizing, Gauthier wants to reformulate this conclusion so that it reads "each person relate his choice of strategy and expectations of other's choices to a set of mutually centroid utility maximizing strategies."

There are potential cases, however, where there are no

---

[18] One might wonder whether these assumptions are necessary for Gauthier. Does he, in fact, require these assumptions or does he merely use them as a device for introducing the centroid utility maximizing response. I raise this question because the centroid utility maximizing response seems to follow too readily from these assumptions.

[19] Morals by Agreement, op cit., p. 68.

centroid utility maximizing strategies. We saw this in the case of Jane and Brian when Jane chose a mixed strategy of assigning a probability of 1/4 to going to the party and 3/4 to staying at home and where Brian's mixed strategy was to assign a probability of 1/2 to going and 1/2 to staying at home. In this case we do have an expected outcome which is in strong equilibrium but we do not have any mutually centroid utility maximizing strategies. If Jane alters her strategy then Brian alters his; if Brian alters his strategy then Jane also alters hers. We can generate a centroid utility maximizing response for each of the participants in this interaction but we cannot generate a mutually centroid utility maximizing response.

Given this example, where Brian assigns an equal probability to going to the party and to staying at home, Jane's centroid utility maximizing response is her mixed strategy of also assigning equal probabilities to going to the party and to staying at home. But if Jane does this then Brian's strategy is not a utility maximizing response to Jane's strategy. In order for Brian's strategy to be a utility maximizing response to this new strategy of Jane's he would have to go to the party. Thus, while we can generate a centroid utility maximizing response for either Jane and Brian we cannot generate one for both. This also illustrates why the outcome is in weak equilibrium rather than strong equilibrium. For the outcome to be in strong equilibrium it would need to be the product of a set of mutually centroid

utility maximizing responses but as we have seen Gauthier was unable to generate such a set, hence, weak equilibrium.

Let us reflect for a moment on what has happened thus far. It will be remembered that Gauthier must formulate a principle of strategic interaction that satisfies the idealizing conditions A, B and C. His first step was to enlarge the scope of choice to include lotteries with actions as prizes. The reason Gauthier enlarged the scope of choice to include lotteries as prizes was so that he could make the move of having Jane and Brian employ mixed strategies in strategic interaction, as opposed to pure strategies. Once they employ mixed strategies, as opposed to pure strategies, then it is possible for them to obtain an equilibrium outcome. An outcome is in equilibrium iff it is product of mutually utility maximizing strategies.

If the outcome is in strong equilibrium then each strategy is the actor's only utility maximizing response to other strategies. If the outcome is in weak equilibrium then there is more than one strategy which is a utility maximizing response to other strategies. If there is more than one strategy which is a utility maximizing response to other strategies then Gauthier must show how one is to choose between strategies. To resolve this potential problem Gauthier introduced the idea of a centroid utility maximizing response where a centroid utility maximizing response is the rational response one makes to the choices one

expects the other to make. This centroid utility maximizing response must be determined by lottery for if it were not one would then be revealing one's preferences for a given outcome which would, in turn, violate the two assumption about preferences being revealed in rational choice and that expected outcomes be equally utility maximizing.

By defining "rational response" in Condition A as a "utility maximizing response" Gauthier wanted to draw two conclusions. One conclusion which he wanted to draw was that "each person should relate his choice of strategy and expectations of other's choices to a set of mutually utility maximizing strategies." This tactic lead to a potential violation of the two assumptions outlined above. In order to avoid violating these two assumptions Gauthier needed to reformulate this conclusion so that it involved a set of mutually centroid utility maximizing strategies, as opposed to a set of mutually utility maximizing strategies.

Having examined the problem where there is no set of mutually centroid utility maximizing responses Gauthier then turns to the case where there is more than one set of mutually centroid utility maximizing responses. If there is more than one set of such responses then the question arises as to how one is to choose among such sets? How does one satisfy condition C of strategically rational choice? To rephrase the question, we can ask, how does one, if there is more than one set of mutually

centroid utility maximizing responses available to choose from, relate one's own choices and expectations to the expectations of the other?

Gauthier suggests that there are two steps involved in circumventing this problem. The first step is to consider the case in which only one set yields an undominated equilibrium outcome. Gauthier offers the following valid argument to show that each person would choose the strategy that belonged to the set yielding the unique undominated equilibrium outcome. That is to say if there is one, and only one, undominated equilibrium outcome then everyone must prefer it to some other outcome.

> An equilibrium outcome is dominated if everyone disprefers it to some other equilibrium outcome. Since no outcome dominates itself and domination is a transitive relation, then an equilibrium outcome cannot be dominated only by other dominated equilibrium outcomes; if it is dominated, then there must be an undominated equilibrium dominating it. A unique undominated equilibrium outcome must therefore be preferred by everyone to every other equilibrium outcome.[20]

The second step to circumvent the problem of relating one's choices and expectations to the choices of others, when there is more than one set of mutually centroid utility maximizing responses to choose from, is to consider the case where there is only one set which yields an undominated equilibrium outcome that no one disprefers to any other equilibrium outcome (which is to

---

[20] Morals by Agreement, op cit., p. 71.

say that there is more than one undominated equilibrium outcome but only one which no one disprefers to any other).

Gauthier's argument for this second step is quite simple. Assume two undominated equilibrium outcomes, say X and Y where outcome Y is the outcome which no one disprefers to any other undominated equilibrium outcome. Then three things follow: first it follows that there will be some who are indifferent between X and Y. Second, it follows that there will be some who prefer Y to X and third it follows that there will be no one who prefers X to Y.

We can easily relate this argument to the problem of how one is to relate one's choices and expectations to the expectations of the other. One relates one's own choices and expectations to the undominated equilibrium outcome which is not dispreferred by anyone to any other equilibrium outcome, and we would expect the other party to do the same. Thus each person involved in the interaction would be correspondingly relating their choices and expectations to the expectations of the other.

The next problem faced by Gauthier is to resolve the problem faced by Victor and Valerie assuming that they cannot form a determinate expectation about the other's choice of strategy.[21] The reason they can't form a determinate expectation about the

---

[21] Morals by Agreement, op cit., pp. 73-75.

other's choice of strategy is because they are involved in a situation where there are two outcomes in strong equilibrium. If neither of them are able to form a determinate expectation about the other's choice of strategy then it follows that neither is able to determine his own utility maximizing response. To resolve this problem Gauthier introduces the notion of "maximin utilities." [22]

The essence of maximin utilities is to maximize the least you might expect to get. If both parties to the interaction are to satisfy the three conditions, A, B, and C, of strategically rational choice then Gauthier thinks that this is one way of doing it. Each considers the utility of the three alternatives afforded to him. Thus Victor would consider the utility to him of his most favored equilibrium (1), the utility to him of Valerie's most favored equilibrium (2/3) and the utility to him of his maximin (1/2). Valerie also considers her three alternatives as well, the utility to her of her most favored equilibrium (1), the utility to her of Victor's most favored equilibrium (1/3) and the utility to her of her maximin (1/3).

Having considered the utilities of these equilibrium outcomes then both Victor and Valerie determine their relative reluctance to concede to the other. Their relative reluctance to concede is

---

[22] Morals by Agreement, op cit., p. 74. A "maximin utility" is the result of acting so that one maximizes the minimum utility one might receive given the situation one finds himself in.

the ratio of their costs in either conceding or settling for their maximin. Thus Victor's relative reluctance to concede is 2/3 and Valerie's relative reluctance to concede is 1. [23] Gauthier then applies Zeuthen's principle; this principle states that "the person whose ratio between cost of concession and cost of deadlock is less must rationally concede to the other." [24] In this case Victor must concede because his ratio is 2/3 while Valerie's is 1. Thus once again Gauthier has found grounds for satisfying condition C, i.e., relating our choices and expectations to the expectations of the other.

The next step Gauthier makes is to challenge the idea of defining rational response in condition A as utility maximizing. He now suggests that we might be better off if we replace utility maximizing response with optimizing response. To say that an

---

[23] One determines their ratio by converting the fraction to a common denominator and then extracting the two numerators. Thus in Victor's case we have the following calculations:

1 - 2/3 = 1/3
1 - 1/2 = 1/2

1/3 = 2/6
1/2 = 3/6

Taking the numerators from 2/6 and 3/6 we get a ratio of 2/3.

Valerie's calculations are the same:

1 - 1/3 = 2/3
1 - 1/3 = 2/3

Taking the numerators from 2/3 and 2/3 we get a ratio of 2/2 or 1.

[24] Morals by Agreement, op. cit., pp. 74-75.

expected outcome is optimizing is just to say "there is no possible outcome affording some person a greater utility and no person a lesser utility." [25] The reason why one might want to consider changing from utility maximizing response to optimizing response is that all optimizing responses are utility maximizing (in the long run) but not all utility maximizing responses are optimal. Then, the argument continues, one can better maximize one's utility, overall, by adopting an optimizing response over a utility maximizing one.

Furthermore, when considering the real world, we find the natural condition of mankind, so accurately described by Hobbes and Hume, to be an "irrational condition." [26] Thus we are, unescapably, led to PD type situations, which in turn leads to outcomes which are not optimal if we define rational response as utility maximizing. We can resolve these dilemmas by exchanging a utility maximizing response in favour of an optimizing response, and, it is rationally required, Gauthier argues, for agents defined as utility maximizers, to do so.

This is the essence of Gauthier's formal apparatus for resolving the conceptual problems associated with strategic interaction. The arguments for complying with the agreements one

---

[25] Morals by Agreement, op cit., p. 76. This is the Pareto-optimality which we discussed in Chapter 1.

[26] Morals By Agreement; op cit., p. 81.

has made are to be found in the two proceeding chapters.

Appendix B

```
**********************************************************************
      Eureka: The Solver, Version 1.0


      Friday May 27, 1988, 2:18 pm.
      Name of input file: B:\CSSIMPLE.
**********************************************************************

            Simple Society: only CM's & only C-S's

     This program finds a value for both c and nc that will prove
that CS > CM. In this case it is rational to choose a C-S
disposition rather than a CM disposition given the assigned
values of d-1, e-0, k-10 and x-1.

                        Variable Assignment

d  - defection
c  - cooperation
nc - non-cooperation
e  - exploitation
k  - number of cooperative opportunities
x  - number of occasions a C-S is in a defect mode

                        Value Assignment

d = 1
c > 0: c < 1: c > nc
nc > 0: nc < 1
e = 0
k = 10
x = 1

      Problem:

Find values of c and nc to show that CS > CM

CS > CM

                            Expressions

CM = k*c

CS = ((k-x)*((k-x)/k)*c) + ((k-x)*(x/k)*e) + (x*((k-x)/k)*d) +
  (x*(x/k)*nc)
```

```
*****************************************************************
      Eureka: The Solver, Version 1.0

      Friday May 27, 1988, 2:18 pm.
      Name of input file: B:\CSSIMPLE.
*****************************************************************


      Solution:

Variables      Values

c          -      .35514066

CM         -      3.5514066

CS         -      3.7880388

d          -      1.0000000

e          -      .00000000

k          -      10.000000

nc         -      .11399484

x          -      1.0000000



*****************************************************************
```

```
**********************************************************************
      Eureka: The Solver, Version 1.0

      Friday May 27, 1988, 2:28 pm.
      Name of input file: B:\SIMPLECM
**********************************************************************
```

Simple Society: only CM's & only C-S's

This program finds a value for both c and nc that will prove that CM > CS. In this case it is rational to choose a CM disposition rather than a C-S disposition given the assigned values of d-1, e-0, k-10 and x-1.

Variable Assignment

d - defection
c - cooperation
nc - non-cooperation
e - exploitation
k - number of cooperative opportunities
x - number of occasions a C-S is in a defect mode

Value Assignment

d = 1
c > 0: c < 1: c > nc
nc > 0: nc < 1
e = 0
k = 10
x = 1

Problem:

Find value of c and nc to show that CM > CS

CM > CS

Expressions

```
CM = k*c

CS = ((k-x)*((k-x)/k)*c) + ((k-x)*(x/k)*e) + (x*((k-x)/k)*d) +
  (x*(x/k)*nc)
```

```
**************************************************************************
      Eureka: The Solver, Version 1.0

      Friday May 27, 1988, 2:28 pm.
      Name of input file: B:\SIMPLECM

**************************************************************************

      Solution:

Variables     Values

c             =      .98215072

CM            =      9.8215072

CS            =      8.9503871

d             =      1.0000000

e             =      .00000000

k             =      10.000000

nc            =      .94966279

x             =      1.0000000



**************************************************************************
```

```
**************************************************************************
      Eureka: The Solver, Version 1.0

      Friday May 27, 1988, 2:54 pm.
      Name of input file: B:\COMPLXCS.

**************************************************************************
```

Complex Society of only CM's and only C-S's

This program finds a value for c, nc and r1 that will prove that CS > CM. In this case it is rational to choose a C-S disposition rather than a CM disposition given the assigned values of d=1, e=0, k=10 and x=1.

Variable Assignment

```
d       - defection
c       - cooperation
nc      - non-cooperation
e       - exploitation
k       - number of cooperative opportunities
x       - number of occasions a C-S is in a defect mode
r1      - ratio of CM's in population
(1-r1)  - ratio of C-S's in population
```

Value Assignment

```
d = 1
c > 0: c < 1: c > nc
nc > 0: nc < 1
e = 0
k = 10
x = 1
r1 < 1
```

Problem:

Find value of c, nc, and r1 to show that CS > CM

CS > CM

Expressions

```
CM = (k*(1-r1)*(x/k)*e) + (k*r1*c) + (k*(1-r1)*((k-x)/k)*c)


CS     =     ((k-x)*(x/k)*(1-r1)*e)     +     ((k-x)*r1*c)     +
     ((k-x)*(1-r1)*((k-x)/k)) + (x*r1*d) + (x*(1-r1)*((k-x)/k)*d)
     + (x*(1-r1)*(x/k)*nc)
```

```
**********************************************************************
      Eureka: The Solver, Version 1.0

      Friday May 27, 1988, 2:54 pm.
      Name of input file: B:\COMPLXCS.

**********************************************************************

      Solution:

Variables     Values

c             -      .98965239

CM            -      9.8436209

CS            -      9.8591360

d             -      1.0000000

e             -      .00000000

k             -      10.000000

nc            -      .97701451

rl            -      .94654389

x             -      1.0000000



**********************************************************************
```